

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

in on: February 29, 2004, 14:51:24 ; Search time 32.998 Seconds  
(without alignments)  
3180.293 Million cell updates/sec

itle: US-09-941-947A-24

irect score: 2598

equences: 1 MNSNDNRQVIVIGAGLGGLS.....NPMVTLGSQLVRDKIVADLQ 497

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 809742 seqs, 21153259 residues

cal number of hits satisfying chosen parameters: 809742

imum DB seq length: 0

imum DB seq length: 2000000000

et-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

itabase : Published Applications AA:

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	2598	100.0	497	9	US-09-934-903-18
2	2598	100.0	497	9	US-09-934-868-78
3	2598	100.0	497	10	US-09-941-947A-24
4	2598	100.0	497	14	US-10-358-917-8
5	1289	49.6	497	14	US-10-358-917-14
6	1039	40.0	439	9	US-09-815-242-5696
7	809	31.1	517	15	US-10-369-493-19519
8	687	26.4	492	10	US-09-941-947A-32
9	687	26.4	492	14	US-10-218-118-8
10	674	25.9	491	15	US-10-369-493-234
11	674	25.9	502	14	US-10-358-917-12
12	662.5	25.5	512	15	US-10-369-493-18644
13	654	25.2	499	15	US-10-369-493-18983
14	645	24.8	498	15	US-10-369-493-20438
15	637	24.5	494	9	US-09-547-267-5

16	637	24.5	494	10	US-09-920-923-4	Sequence 4, Appli
17	628	24.2	494	15	US-10-369-493-8951	Sequence 8951, Ap
18	619.5	23.8	498	15	US-10-369-493-19248	Sequence 19248, A
19	617	23.7	506	15	US-10-369-493-17830	Sequence 17830, A
20	607.5	23.4	511	9	US-09-934-903-16	Sequence 16, Appl
21	607.5	23.4	511	9	US-09-934-868-76	Sequence 76, Appl
22	607.5	23.4	511	10	US-09-941-947A-22	Sequence 22, Appl
23	607.5	23.4	511	14	US-10-358-917-4	Sequence 4, Appli
24	596	22.9	290	9	US-09-939-980-462	Sequence 462, App
25	590.5	22.7	514	15	US-10-369-493-21541	Sequence 21541, A
26	586	22.6	618	15	US-10-369-493-3769	Sequence 3769, Ap
27	576	22.2	498	15	US-10-369-493-10579	Sequence 10579, A
28	568.5	21.9	536	15	US-10-369-493-18646	Sequence 18646, A
29	564.5	21.7	496	15	US-10-369-493-20440	Sequence 20440, A
30	551	21.2	518	15	US-10-369-493-7750	Sequence 7750, Ap
31	550.5	21.2	495	15	US-10-369-493-7895	Sequence 7895, Ap
32	544	20.9	548	15	US-10-369-493-540	Sequence 540, App
33	536	20.6	506	15	US-10-369-493-8142	Sequence 8142, Ap
34	529.5	20.4	513	14	US-10-156-761-8562	Sequence 8562, Ap
35	522.5	20.1	1268	15	US-10-438-784-3	Sequence 3, Appli
36	499.5	19.2	498	15	US-10-369-493-10196	Sequence 10196, A
37	475	18.3	548	9	US-09-738-626-4194	Sequence 4194, Ap
38	469.5	18.1	469	15	US-10-438-784-5	Sequence 5, Appli
39	458	17.6	485	15	US-10-369-493-19518	Sequence 19518, A
40	440.5	17.0	404	15	US-10-369-493-19547	Sequence 19547, A
41	440	16.9	497	14	US-10-156-761-13629	Sequence 13629, A
42	411	15.8	530	14	US-10-128-713A-18	Sequence 18, Appl
43	386	14.9	498	15	US-10-369-493-17259	Sequence 17259, A
44	357	13.7	492	15	US-10-369-493-19911	Sequence 19911, A
45	347	13.4	494	15	US-10-369-493-18944	Sequence 18944, A

## ALIGNMENTS

### RESULT 1

US-09-934-903-18  
Sequence 18, Application US/09934903  
Patent No. US20020102690A1  
GENERAL INFORMATION:  
APPLICANT: Koffas, Mattheos  
APPLICANT: Odum, J. Martin  
APPLICANT: Schenzle, Andreas J.  
APPLICANT: No. US20020102690Alton, Kelley C.  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Rouviere, Pierre  
APPLICANT: Picataggio, Stephen  
APPLICANT: Cheng, Qiong  
TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production  
FILE REFERENCE: CL1646 US NA  
CURRENT APPLICATION NUMBER: US/09/934,903  
CURRENT FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/229,907  
PRIOR FILING DATE: September 1, 2001  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 18  
LENGTH: 497  
TYPE: PRT  
ORGANISM: Methylomonas 16a  
FEATURE:  
OTHER INFORMATION: Amino acid sequences encoded by ORF9  
US-09-934-903-18

Query Match	100.0%	Score	2598	DB	9	Length	497
Best Local Similarity	100.0%	Pred. No.	4.6e-251	Mismatches	0	Indels	0
Matches	497	Conservative	0	0	Gaps	0	0
Cy	1	MNSNDNRQVIVIGAGLGGLS	SAISL	ATAGFS	VQLTEKNDKVGGLNIMTKGFTFDL	GPS	60
Db	1	MNSNDNRQVIVIGAGLGGLS	SAISL	ATAGFS	VQLTEKNDKVGGLNIMTKGFTFDL	GPS	60
Qy	61	ILTMPHIPEALFTGAGKNMADYVQIQKVEPHWRNPFEDGSVIDLCEDAETQRRELDKLP					120

Db 61 ILTMPIFEALFTGAGKKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETORRELDKLG 120  
QY 121 GTYAQORFLDYSKULTTETAGYFAKLDGFDNLLKFYGLRSLISFDVFRSMDQVRR 180  
Db 121 GTYAQORFLDYSKULTTETAGYFAKLDGFDNLLKFYGLRSLISFDVFRSMDQVRR 180  
QY 181 FISDPKLVILNFIKYVGGSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240  
Db 181 FISDPKLVILNFIKYVGGSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240  
QY 241 ELGVIRLDAEVSLEOKQDGRACAVKLANGVDLPADIWVSNMEVI PAMEKLLRSPASELK 300  
Db 241 ELGVIRLDAEVSLEOKQDGRACAVKLANGVDLPADIWVSNMEVI PAMEKLLRSPASELK 300  
QY 301 KQORPEPSCGLVHLGVDRLYPOLAHNFYSDHPREHFDVAFKSHRLSDDPTIYLVP 360  
Db 301 KQORPEPSCGLVHLGVDRLYPOLAHNFYSDHPREHFDVAFKSHRLSDDPTIYLVP 360  
QY 361 CKTDPAPAGCEIIKILPHILDPDKLLTAEDYSALRERVLVKLERMGLTDLRQHIVT 420  
Db 361 CKTDPAPAGCEIIKILPHILDPDKLLTAEDYSALRERVLVKLERMGLTDLRQHIVT 420  
QY 421 EYWTPLDIQAKYYSNOGSIYGVADRFPKNGFKAPQRSSELSNLYFVGGSVNPGGMPM 480  
Db 421 EYWTPLDIQAKYYSNOGSIYGVADRFPKNGFKAPQRSSELSNLYFVGGSVNPGGMPM 480  
QY 481 VTLSGQLVRDKIVADLQ 497  
Db 481 VTLSGQLVRDKIVADLQ 497

## RESULT 2

US-09-934-868-78  
; Sequence 78, Application US/09934868  
; Patent No. US20020137190A1  
; GENERAL INFORMATION:  
; APPLICANT: Koffas, James M  
; APPLICANT: Odom, James M  
; APPLICANT: Schenzle, Andreas J  
; TITLE OF INVENTION: IDENTIFYING METHANOTROPHIC BACTERIAL STRAIN  
; FILE REFERENCE: CL1596 US NA  
; CURRENT APPLICATION NUMBER: US/09/934,868  
; CURRENT FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/229,858  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 81  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 78  
; LENGTH: 497  
; TYPE: PRT  
; ORGANISM: Methylobionas 16a  
; FEATURE:  
; OTHER INFORMATION: Amino acid sequences encoded by CRTW2  
US-09-934-868-78

Query Match 100.0%; Score 2598; DB 9; Length 497;  
Best Local Similarity 100.0%; Pred. No. 4.6e-251;  
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNSNDNRVIVIGAGLGGLSAASISLATAGFSVOLIEKNDKVGKLNIMTKDGTFFDLGSP 60  
Db 1 MNSNDNRVIVIGAGLGGLSAASISLATAGFSVOLIEKNDKVGKLNIMTKDGTFFDLGSP 60  
QY 61 ILTMPIFEALFTGAGKKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETORRELDKLG 120  
Db 61 ILTMPIFEALFTGAGKKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETORRELDKLG 120  
QY 121 GTYAQORFLDYSKULTTETAGYFAKLDGFDNLLKFYGLRSLISFDVFRSMDQVRR 180  
Db 121 GTYAQORFLDYSKULTTETAGYFAKLDGFDNLLKFYGLRSLISFDVFRSMDQVRR 180  
QY 181 FISDPKLVILNFIKYVGGSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240

Db 181 FISDPKLVILNFIKYVGGSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240  
QY 241 ELGVIRLDAEVSLEOKQDGRACAVKLANGVDLPADIWVSNMEVI PAMEKLLRSPASELK 300  
Db 241 ELGVIRLDAEVSLEOKQDGRACAVKLANGVDLPADIWVSNMEVI PAMEKLLRSPASELK 300  
QY 301 KQORPEPSCGLVHLGVDRLYPOLAHNFYSDHPREHFDVAFKSHRLSDDPTIYLVP 360  
Db 301 KQORPEPSCGLVHLGVDRLYPOLAHNFYSDHPREHFDVAFKSHRLSDDPTIYLVP 360  
QY 361 CKTDPAPAGCEIIKILPHILDPDKLLTAEDYSALRERVLVKLERMGLTDLRQHIVT 420  
Db 361 CKTDPAPAGCEIIKILPHILDPDKLLTAEDYSALRERVLVKLERMGLTDLRQHIVT 420  
QY 421 EYWTPLDIQAKYYSNOGSIYGVADRFPKNGFKAPQRSSELSNLYFVGGSVNPGGMPM 480  
Db 421 EYWTPLDIQAKYYSNOGSIYGVADRFPKNGFKAPQRSSELSNLYFVGGSVNPGGMPM 480  
QY 481 VTLSGQLVRDKIVADLQ 497  
Db 481 VTLSGQLVRDKIVADLQ 497

## RESULT 3

US-09-941-947A-24  
; Sequence 24, Application US/09941947A  
; Publication No. US20030003528A1  
; GENERAL INFORMATION:  
; APPLICANT: Brzostowicz, Patricia C.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: DiCosimo, Deana J.  
; APPLICANT: Koffas, Mattheos  
; APPLICANT: Miller, Edward S. Jr.  
; APPLICANT: Odom, J. Martin  
; APPLICANT: Picataggio, Steve  
; APPLICANT: Rouviere, Pierre E.  
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE  
; FILE REFERENCE: CL1903 US NA  
; CURRENT APPLICATION NUMBER: US/09/941,947A  
; CURRENT FILING DATE: 2001-09-01  
; PRIOR APPLICATION NUMBER: 60/229,907  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,858  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 24  
; LENGTH: 497  
; TYPE: PRT  
; ORGANISM: Methylobionas 16a  
US-09-941-947A-24

Query Match 100.0%; Score 2598; DB 10; Length 497;  
Best Local Similarity 100.0%; Pred. No. 4.6e-251;  
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MNSNDNRVIVIGAGLGGLSAASISLATAGFSVOLIEKNDKVGKLNIMTKDGTFFDLGSP 60  
Db 1 MNSNDNRVIVIGAGLGGLSAASISLATAGFSVOLIEKNDKVGKLNIMTKDGTFFDLGSP 60  
QY 61 ILTMPIFEALFTGAGKKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETORRELDKLG 120  
Db 61 ILTMPIFEALFTGAGKKNADYVQIQKVEPHRNFFEDGSDVLDCEDAETORRELDKLG 120  
QY 121 GTYAQORFLDYSKULTTETAGYFAKLDGFDNLLKFYGLRSLISFDVFRSMDQVRR 180  
Db 121 GTYAQORFLDYSKULTTETAGYFAKLDGFDNLLKFYGLRSLISFDVFRSMDQVRR 180  
QY 181 FISDPKLVILNFIKYVGGSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240  
Db 181 FISDPKLVILNFIKYVGGSSPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLV 240

241 ELGVEIRLDAEVSBEIQODGRACAVKLANGDVLADIVSNMEVIPAMEKLSPASELK 300  
|||||  
241 ELGVEIRLDAEVSBEIQODGRACAVKLANGDVLADIVSNMEVIPAMEKLSPASELK 300  
|||||  
301 KQRFEPSCSLVHLGLVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDPTIYLVA 360  
|||||  
301 KQRFEPSCSLVHLGLVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDPTIYLVA 360  
|||||  
361 CKTDPAPACCEIIKILPHILDPDKLTADYSALRERVLVKLRMGLTDLRQHIVT 420  
|||||  
361 CKTDPAPACCEIIKILPHILDPDKLTADYSALRERVLVKLRMGLTDLRQHIVT 420  
|||||  
421 EEWTPDLIOAKYYSNOGSIYGVVADRFKNGKAPQRSSELSNLYFVGGSVNPGGMPM 480  
|||||  
421 EEWTPDLIOAKYYSNOGSIYGVVADRFKNGKAPQRSSELSNLYFVGGSVNPGGMPM 480  
|||||  
481 VTLGQVLVDKIVADLQ 497  
|||||  
481 VTLGQVLVDKIVADLQ 497  
|||||

SULT 4  
-10-358-917-8  
Sequence 8, Application US/10358917  
Publication No. US20030182687A1  
GENERAL INFORMATION:  
APPLICANT: Cheng, Qiong  
APPLICANT: No. US20030182687A1ton, Kelley C.  
APPLICANT: Tao, Luan  
TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOID COMPOUNDS  
FILE REFERENCE: CL1929 US NA  
CURRENT APPLICATION NUMBER: US/10/358,917  
CURRENT FILING DATE: 2003-02-05  
PRIOR APPLICATION NUMBER: 60/355,939  
PRIOR FILING DATE: 2002-02-11  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 8  
LENGTH: 497  
TYPE: PRT  
ORGANISM: Methylomonas sp.16a  
-10-358-917-8

Query Match 100.0%; Score 2598; DB 14; Length 497;  
Best Local Similarity 100.0%; Pred. No. 4.6e-251;  
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNSNDNQRVIVIGAGLGLSAAISLATAGPSVQLIEKNDKVGKLNIMTKDGTFFDLGPS 60  
|||||  
1 MNSNDNQRVIVIGAGLGLSAAISLATAGPSVQLIEKNDKVGKLNIMTKDGTFFDLGPS 60  
|||||  
61 ILTMPIFEALFTGAGKNMADYVQIQKVEPHRNFFEDGSDVIDLCEAETORRELDKLP 120  
|||||  
61 ILTMPIFEALFTGAGKNMADYVQIQKVEPHRNFFEDGSDVIDLCEAETORRELDKLP 120  
|||||  
121 GTYAFORFLDYKSLCTETAGYFAKGLDGFNDLLKPYGLRSLSDFDVFRSMDQVRR 180  
|||||  
121 GTYAFORFLDYKSLCTETAGYFAKGLDGFNDLLKPYGLRSLSDFDVFRSMDQVRR 180  
|||||  
181 FTSDPKLVEILNFIKYVGSSPYDAPALMNLPIYQHYGLWYKGMYGMAQAMEKLAV 240  
|||||  
181 FTSDPKLVEILNFIKYVGSSPYDAPALMNLPIYQHYGLWYKGMYGMAQAMEKLAV 240  
|||||  
241 ELGVEIRLDAEVSBEIQODGRACAVKLANGDVLADIVSNMEVIPAMEKLSPASELK 300  
|||||  
241 ELGVEIRLDAEVSBEIQODGRACAVKLANGDVLADIVSNMEVIPAMEKLSPASELK 300  
|||||  
301 KQRFEPSCSLVHLGLVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDPTIYLVA 360  
|||||  
301 KQRFEPSCSLVHLGLVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDPTIYLVA 360  
|||||  
361 CKTDPAPACCEIIKILPHILDPDKLTADYSALRERVLVKLRMGLTDLRQHIVT 420  
|||||

361 CKTDPAPACCEIIKILPHILDPDKLTADYSALRERVLVKLRMGLTDLRQHIVT 420  
|||||  
421 EEWTPDLIOAKYYSNOGSIYGVVADRFKNGKAPQRSSELSNLYFVGGSVNPGGMPM 480  
|||||  
421 EEWTPDLIOAKYYSNOGSIYGVVADRFKNGKAPQRSSELSNLYFVGGSVNPGGMPM 480  
|||||  
481 VTLGQVLVDKIVADLQ 497  
|||||  
481 VTLGQVLVDKIVADLQ 497  
|||||

RESULT 5  
US-10-358-917-14  
Sequence 14, Application US/10358917  
Publication No. US20030182687A1  
GENERAL INFORMATION:  
APPLICANT: Cheng, Qiong  
APPLICANT: No. US20030182687A1ton, Kelley C.  
APPLICANT: Tao, Luan  
TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOID COMPOUNDS  
FILE REFERENCE: CL1929 US NA  
CURRENT APPLICATION NUMBER: US/10/358,917  
CURRENT FILING DATE: 2003-02-05  
PRIOR APPLICATION NUMBER: 60/355,939  
PRIOR FILING DATE: 2002-02-11  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 14  
LENGTH: 497  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-10-358-917-14

Query Match 49.6%; Score 1289; DB 14; Length 497;  
Best Local Similarity 51.9%; Pred. No. 6e-120;  
Matches 254; Conservative 85; Mismatches 148; Indels 2; Gaps 2;

7 QRVIVIGAGLGLSAAISLATAGPSVQLIEKNDKVGKLNIMTKDGTFFDLGPSILTMPH 66  
|||||  
3 KHIIVIGGLGGSIAIRMAQSGSVLSYEQNNHIGGVKVRHESDGFGLGPSILTMPY 62  
|||||  
67 IFEALFTGAGKNMADYVQIQKVEPHRNFFEDGSDVIDLCEAETORRELDKLPFTVAQF 126  
|||||  
63 IFEKLFYSKXQMSDYVTIKRLPHQMSFFPDGTTIDLYEGIKETGQHNAILSKQDIEEL 122  
|||||  
127 QRPDYKSLCTETAGYFAKGLDGFNDLLKPYGLRSLSDFDVFRSMDQVRRFISDPK 186  
|||||  
123 QNYLTYTRIDRITEKGYFNGLDGLTSLQIIKPHGLNALINDYVHTMQQADIKRISNPY 182  
|||||  
187 LVEILNFIKYVGSSPYDAPALMNLPIYQHYGLWYKGMYGMAQAMEKLAVELGVSI 246  
|||||  
183 LRQMLGYFIKYVGSSPYDAPALMNLPIYQHYGLWYKGMYGMAQAMEKLAVELGVSI 242  
|||||  
247 RLDAVSBEIQODGRACAVKLANGDVLADIVSNMEVIPAMEKLSPASELKMMOR-F 305  
|||||  
243 HPGARVDNIKITQYRVTVGLTGTGCFVKADYIIISNMEVITYKYLHLDTORLNKLRREF 302  
|||||  
306 EPCSGGLVHLGLVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDPTIYLVAFCPTDP 365  
|||||  
303 EPASSGYVHLGVACQYQOLAHNFFFTENAYLVAVQVFEKVLPPDPTIYLVTNKTGH 362  
|||||  
366 AQAPAGCEIIKILPHILDPDKLTADYSALRERVLVKLRMGLTDLRQHIVTEYVWT 425  
|||||  
363 TQAPVGVENIKVLPHPYIQ-DQPTTDEYAFKDKLDEKMGTLDRKHIIYEOWMT 421  
|||||  
426 PLDIOAKYYSNOGSIYGVVADRFKNGKAPQRSSELSNLYFVGGSVNPGGMPMVTLSG 485  
|||||  
422 PEDISKTRSRNGALYGVVADKKNKGKGFPPKESQYFENLYFVGGSVNPGGMPMVTLSG 481  
|||||  
486 QLVRDKIVA 494  
|||||  
482 QQVADKINA 490  
|||||

## RESULT 6

IS-09-815-242-5696  
Sequence 5696, Application US/09815242  
Patent No. US20020061569A1

## GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert  
APPLICANT: Orlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5696

LENGTH: 439

TYPE: PRT

ORGANISM: Staphylococcus aureus

IS-09-815-242-5696

Query Match 40.0%; Score 1039; DB 9; Length 439;

Best Local Similarity 48.7%; Pred. No. 5.4e-95;

Matches 205; Conservative 80; Mismatches 134; Indels 2; Gaps 2;

2y 7 QRIVIGAGLGLSAAISLATAGSVOLIEKNKVGKLNIMTKDGTFTDLGPSILTMH 66

3 KHIIVIGGLGSIISAIRMAQSGSVSLYEQNHIGKVNHRHSDGFGFDLGPSILTMPI 62

67 IFEALFTGAGKNMADYVQIKVEPHRNFFEDSGVIDLCEDAETQRELDKLGPGTYAQF 126

63 IFEKLFYSSKKQMSDYTIKRLPHQWSFFPDGTTIDLYEGIKETQHAILSNKDIEL 122

127 QRFLDYSKNLCTETAGYFAKGLDGLGFWDLKFGPLRSLSLSDVFRSMQGVRRFISDFK 186

123 QNLYNTRRIDRIETKGVNGLDLSQIKFKGPNALINDYVHTWQQAIDKRISNPF 182

187 LVEILNFIKYVSSPDAPALMNLPIYQHYGLVYKGMVGMAGMEKLAELVGEI 246

183 LRQMLGPIKYVSSSDYDAVLSNLFHMQOGLWYVEGIIHLNALKLAREGVTI 242

247 RLDAEVESEIQKQGRACAVKLANGDVLPAIVVSNMEVIPAMEKLLRSPASSLKQMR-F 305

243 HTGTRVDNIKTYQRVTVGLRDLTGEFVKADVIISNMEVITYKYLLHLGQLRNLKLEREF 302

306 BPSCSGLVHLGVDRLLPOLAHNFFVSDPREHFDVAFKSHLSDDPTIYVAPCKTDP 365

303 EPASSGVYMLGVACQYPOLAHNFFFTENAYLNYYQVFEHKVLPDDPTIYLVNTKTDH 362

366 AQAPAGCEIILKILPHIPLDOKLLTAEDYSALRERVLKLERMGLTDLRQHIIVTEEYT 425

363 TQAPVGVENIKVLPHIPIYQ-DQPTFTEDAKFDKILDKLERMGLTDLRKYYIYEDVNI 421

OY 426 P 426

Db 422 P 422

## RESULT 7

US-10-369-493-19519

Sequence 19519, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 19519

LENGTH: 517

TYPE: PRT

ORGANISM: Myxococcus xanthus

US-10-369-493-19519

Query Match 31.1%; Score 809; DB 15; Length 517;

Best Local Similarity 34.9%; Pred. No. 7.5e-72;

Matches 175; Conservative 100; Mismatches 204; Indels 22; Gaps 7;

OY 3 SNDQNRVIVIGAGLGLSAAISLATAGSVOLIEKNKVGKLNIMTKDGTFTDLGPSIL 62

Db 4 STQGRIVVVGAGVGGGLAAARLAHQGFQVQFEKTKQPGRCNRLQVQDFTWDLGPTIV 63

63 TMPIFEALFTGAGKNMADYVQIKVEPHRNFFEDSGVIDLCEDAETQRELDKLGPGT 122

64 LMPEVFETFRVNGRIEDVLTLLRCDFNVYVHFRDGSVDVTFSELCAWGRELERVEPGS 123

123 YAQQRFLDYSKNLCTETAGYFAKGLDGLGFWDLKFGPLRSLSLSDVFRSMQGVRRFISDFK 173

124 YARYLAFLAQGR-----VQRTSLDHL--VGRNYAGLEDYLSPRVLARIFQVRAHRR 173

174 MDQGVRRFISDPKIVELNFIKYVSSPDAPALMNLPIYQHYGLVYKGMVGMAGMEKLAELVGEI 233

174 MYADVSFFODERLRAMATQTMYLGVSPYASVAVGLLPFTTELGVIGWEPKGLVAIPO 233

234 AMEKLAELVGEIRLDAEVESEIQKQGRACAVKLANGDVLPAIVVSNMEVIPAMEKLLR 293

234 ALERLARBEQVRFEYGAVERILTDGGRTRGVLEGEVVEADAVLCNADLPYAEKLLD 293

294 SPASELKQMRFPBSCSGLVHLGVDRLLPOLAHNFFVSDPREHFDVAFKSHRLSDDP 353

294 PKATTLRKEKELRYTSSGYMLYLGKGRYPDELHLNVPVFGDYKSGFDDIFERFRVPEDP 353

354 TIYLVAPCKTDPAQAPAGCEIILKILPHIPLDOKLLTAEDYSALRERVLKLERMGLTD 413

354 SFYVNAPTRIDASLAPGSKDALVYLVVPVPHOHFDLQWKVEG-PKVRAKFPARMABELGPPS 412

414 LRQHIIVTEEWTPLDIOAKYKSNQSIYGVVADRFKMLG-PKAPQRSSELSNLYFVGSSV 472

413 LESDIEVERVFTPDWAGTFNLARGSAFG-LSONFTQIGPFRPSNQARVKNLFFVGAST 471

473 NPGGMPMTVLSGOLVRDKIV 493

472 QPGLGLFTVLISARLVTERLM 492

## RESULT 8

US-09-941-947A-32

Sequence 32, Application US/09941947A



Publication No. US20030030528A1  
GENERAL INFORMATION:

APPLICANT: Brzostowicz, Patricia C.  
APPLICANT: Cheng, Qiong  
APPLICANT: Dicossimo, Deana J.  
APPLICANT: Koffas, Mattheos  
APPLICANT: Miller, Edward S. Jr.  
APPLICANT: Odom, J. Martin  
APPLICANT: Picataggio, Steve  
APPLICANT: Rouviere, Pierre E.

TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE

FILE REFERENCE: CL1903 US NA

CURRENT APPLICATION NUMBER: US/09/941,947A

CURRENT FILING DATE: 2001-09-01

PRIOR APPLICATION NUMBER: 60/229,907

PRIOR FILING DATE: 2000-09-01

PRIOR APPLICATION NUMBER: 60/229,858

PRIOR FILING DATE: 2000-09-01

NUMBER OF SEQ ID NOS: 60

SOFTWARE: Microsoft Office 97

SEQ ID NO 32

LENGTH: 492

TYPE: PR1

ORGANISM: Pantoea stewartii

S-09-941-947A-32

Query Match 26.4%; Score 687; DB 10; Length 492;

Best Local Similarity 33.3%; Pred. No. 1.2e-59;

Matches 164; Conservative 87; Mismatches 229; Indels 12; Gaps 4;

Y 11 VIGAGLGLSAAISLATAGFSVQLIEKNKVGKLNIMTKDGTFTDLGSIILTMHIFEA 70

b 6 VIGAGPGGLALAIRLOAGIPVLLLRQKPGGRAYVYQEQFTFDAGTIVTDSAAIEE 65

Y 71 LFTGAGKXWADYVQIQKVEPHWNTFFEDGSVIDLCEAETQRELDKLGPGTVAQORFL 130

b 66 LFALAGKQLADYVELLPVTPFYRLCWESKGVNFYNDQALEAQIQOENPRDVAGYRAFL 125

Y 131 DYSKNLCTETEAGYFAKGLDGFMDLLKFTGPLES---LLSFDVFRSMQGVRRFISDPKL 187

b 126 DYGRAVNE---GYLKLGTVPF---LSFKDMLRAAPQLAKLQAWRSVYSKVAGYIEDEHL 179

Y 188 VEILNFIKYVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELGVEIR 247

b 180 RQAFPSHLLVGNPNPATSIYTLIHALLEREMGVWPPRGGTGALVNGMIKLPQDLGGEV 239

Y 248 LDAEVSIEIQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKILR---SPASELKKMQR 304

b 240 LNAVSHMETVGDKIQAQVLEDGRRFETCAVASNADVHTYRDLLSQHPAAAKQAKLQS 299

Y 305 FEPSCGLVHLGVDRILYPOLAHNFYSDHPREHFDVAFKSHRLSDDDPTIYLVAECKTD 364

b 300 KRMNSLFLVLYFGLNHHHDLQAHHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPCVD 359

Y 365 PAQAPACGEIILKILPHILPDKILTAEDYSALRERVVLKERMGLTDLRQHIIVTEBYM 424

b 360 PSLAPECCGSYYVLAFFVPHLGTANLDWAVEGPRLRDRIEFDYLEQHYMPGLRSQVTHRMF 419

Y 425 TPDIQAKYYSNCGSIYGVVADRFKNLGFKAPORSSELSNLYFVGSVNPFGGMPMTLS 484

b 420 TPDFRDELNAWQGSAPSVPEPILTSQAWFRPHNRDKHIDNLYLVGAGTTPGAGIPGVIGS 479

Y 485 GOLVRDKIVADL 496

b 480 AKATAGLMLSDL 491

RESULT 9

US-10-218-118-8

Sequence 8, Application US/10218118

Publication No. US20030148319A1

GENERAL INFORMATION:

APPLICANT: Brzostowicz, Patricia

APPLICANT: Rouviere, Pierre  
APPLICANT: Picataggio, Stephen  
APPLICANT: Cheng, Qiong

TITLE OF INVENTION: Genes Encoding Carotenoid Compounds

FILE REFERENCE: CL1976 US NA

CURRENT APPLICATION NUMBER: US/10/218,118

CURRENT FILING DATE: 2002-08-13

PRIOR APPLICATION NUMBER: 60/312,646

PRIOR FILING DATE: 2001-08-15

NUMBER OF SEQ ID NOS: 14

SOFTWARE: Microsoft Office 97

SEQ ID NO 8

LENGTH: 492

TYPE: PR1

ORGANISM: Pantoea stewartii

US-10-218-118-8

Query Match 26.4%; Score 687; DB 14; Length 492;

Best Local Similarity 33.3%; Pred. No. 1.2e-59;

Matches 164; Conservative 87; Mismatches 229; Indels 12; Gaps 4;

Y 11 VIGAGLGLSAAISLATAGFSVQLIEKNKVGKLNIMTKDGTFTDLGSIILTMHIFEA 70

b 6 VIGAGPGGLALAIRLOAGIPVLLLRQKPGGRAYVYQEQFTFDAGTIVTDSAAIEE 65

Y 71 LFTGAGKXWADYVQIQKVEPHWNTFFEDGSVIDLCEAETQRELDKLGPGTVAQORFL 130

b 66 LFALAGKQLADYVELLPVTPFYRLCWESKGVNFYNDQALEAQIQOENPRDVAGYRAFL 125

Y 131 DYSKNLCTETEAGYFAKGLDGFMDLLKFTGPLES---LLSFDVFRSMQGVRRFISDPKL 187

b 126 DYGRAVNE---GYLKLGTVPF---LSFKDMLRAAPQLAKLQAWRSVYSKVAGYIEDEHL 179

Y 188 VEILNFIKYVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELGVEIR 247

b 180 RQAFPSHLLVGNPNPATSIYTLIHALLEREMGVWPPRGGTGALVNGMIKLPQDLGGEV 239

Y 248 LDAEVSIEIQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKILR---SPASELKKMQR 304

b 240 LNAVSHMETVGDKIQAQVLEDGRRFETCAVASNADVHTYRDLLSQHPAAAKQAKLQS 299

Y 305 FEPSCGLVHLGVDRILYPOLAHNFYSDHPREHFDVAFKSHRLSDDDPTIYLVAECKTD 364

b 300 KRMNSLFLVLYFGLNHHHDLQAHHTVCFGRYRELHIEIFNHDGLAEDPSLYLHAPCVD 359

Y 365 PAQAPACGEIILKILPHILPDKILTAEDYSALRERVVLKERMGLTDLRQHIIVTEBYM 424

b 360 PSLAPECCGSYYVLAFFVPHLGTANLDWAVEGPRLRDRIEFDYLEQHYMPGLRSQVTHRMF 419

Y 425 TPDIQAKYYSNCGSIYGVVADRFKNLGFKAPORSSELSNLYFVGSVNPFGGMPMTLS 484

b 420 TPDFRDELNAWQGSAPSVPEPILTSQAWFRPHNRDKHIDNLYLVGAGTTPGAGIPGVIGS 479

Y 485 GOLVRDKIVADL 496

b 480 AKATAGLMLSDL 491

RESULT 10

US-10-369-493-234

Sequence 234, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 234

LENGTH: 491

TYPE: PRT

ORGANISM: Xenorhabdus nematophilus

US-10-369-493-234

Query Match 25.9%; Score 674; DB 15; Length 491;

Best Local Similarity 29.2%; Pred. No. 2.3e-58;

Matches 145; Conservative 111; Mismatches 225; Indels 16; Gaps 3;

8 RVIVIGAGLGLSAAISLATAGFSVOLIEKNKDVGGKLNIMTKDGTFFDLGSPILTMPIH 67

3 RIPVIGSGFGIALAIRLQANIPITLLDRKRGGRAYVYQDKGTFDAGPVTIDPSA 62

68 FEALFTGAGKNMADYVQIQVBEHWRNFEDGSDVIDLCEDAETQRRELKLGPGTYAQFQ 127

63 IEELFLLAGKMSDYIELLPISFPYRLQMSGEIFDYGGQVKLEKQIARFNPGRVQGYR 122

128 RFLDYSKNLCTETAGYFAKGLDGFMDLLKPYGLRSLLSFDFRSMDCQVRRFISDPKL 187

123 EFIRYSKAV---FTGVLYKGMSPFLSFRSMWCAMPSLVQLSAKMSVYGVSKIKNEHL 179

188 VEILNLYFIKYVGGSPYDAPALMNLPIYQHYGLWYKGMVYKGMVYKGMVYKGMVYKGMV 247

180 RQAFSPHSLIIGNPFTTSYITLIHLREGGVWPRGOTGALVNLVQLFIDIGGVCK 239

248 LDAEVSIEIQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLL-----RSPASBL 299

240 LNTQVESFINDNQTAKLNDQREIYCKAASNAADVHTYEKLGSHCVGRSDSLKE 299

300 KMQRPFPSCGLVHLGVRLYQPLAHNPFYSDHPREHDAVFXSHRLSDOPTIYLVVA 359

300 KCM-----SNSLFVYFGLNKLHQLAHTTCFGRNYKELLDDIFEGNSIPGDFSYLHS 354

360 PKTDPRAQAPAGCEIILPHIPLHDPKLLTAEDYSALAEVRLVLERMGLTDLRCHIV 419

355 PSVTDPTLAPPCASYYVLPVPHLGNAGLNWLEESPLRDRIFSYLEKYYIPCLRDQLV 414

420 TEYWTPLDIQAKYNSQSGIYGVVADRFRNKGFKAPQRSSELSNLYFVGGSVNPGGMP 479

415 TEKIFTDFDELNAYHGSFAPIELLQSAWPRPHNRSDIENLYIVGAGTHPGAGIP 474

480 MVLSSGQLVRDKIVADL 496

475 GVLASAKITANLMTISDL 491

RESULT 11

US-10-358-917-12

Sequence 12, Application US/10358917

Publication No. US20030182687A1

GENERAL INFORMATION:

APPLICANT: Cheng, Qiong

APPLICANT: No. US20030182687A1, Kelton, Kelley C.

APPLICANT: Tao, Luan

TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOID COMPOUNDS

FILE REFERENCE: CL1929 US NA

CURRENT APPLICATION NUMBER: US/10/358,917

CURRENT FILING DATE: 2003-02-05

PRIOR APPLICATION NUMBER: 60/355,939

PRIOR FILING DATE: 2002-02-11

NUMBER OF SEQ ID NOS: 28

SOFTWARE: Microsoft Office 97

SEQ ID NO 12

LENGTH: 502

TYPE: PRT

ORGANISM: Staphylococcus aureus

US-10-358-917-12

Query Match 25.9%; Score 672; DB 14; Length 502;

Best Local Similarity 31.5%; Pred. No. 3.8e-58;

Matches 162; Conservative 120; Mismatches 191; Indels 42; Gaps 14;

QY 8 RVIVIGAGLGLSAAISLATAGFSVOLIEKNKDVGGKLNIMTKDGTFFDLGSPILTMPIH 67

Db 2 KIAVIGAGVTGLAARAAIASQGEHVIIFKKNVVGGMNQLKDGFTFDGFTVIVMPDV 61

QY 68 FEALFTGAGKNMADYVQIQVBEHWRNFEDGSDVIDLCEDAETQRRELKLGPGTYAQFQ 127

Db 62 YKDVFTACGKNYEDYIELRQLRYIVDYFQDHDRIITVPTDLAELQOMLESIEPGSTHGM 121

QY 128 RFLD--YSKNLCTETAGYF-----AKGLDGFMDLLKPYGLRSLLSFDFVFRSM---DQG 177

Db 122 SFLTDVYKK---VEIARRYFLERTYRPSD-----FTN-MTSLVQGAKLKTLNADQL 170

QY 178 VRRFISDFKLVEILNLYFIKYVGGSPYDAPALMNLPIYQHYGLWYKGMVYKGMVYKGMVYKGMV 237

Db 171 IERYIDNEKIQLLAPQTLYGIDPKRGPSLYSIIPMIEMMGVHFIKGMVYKGMVYKGMVYKGMV 230

QY 238 LAVELGVBEIRLDAEVSF--IQKQGRACAVKLANGDVLPAIVVSNMEVIPAMEKLLRSP 295

Db 231 LNKDLGVNIELNABIEGIIIDPKFKRADAIVK--NGDIRKFKILCTADPSPVAESLMPDF 289

QY 296 ASEELK---KMQRPFPSCGLVHLGVLD--RLYPQLAHNPFYSDHPREHDAVFXSHRLS 350

Db 290 APIKKYPPHKLADLDYSCAPLAVIGIDIDVTQVRLHNVPISDDFRGNEIEIFEG-RLS 348

QY 351 DDPIYIVAPCKTDPAPAGCEIILPHIPLHDPKLLTAEDYSALAEVRLV 403

Db 349 YDPSIYVYVFAVADKSLAPEGKTIYVIMPTPELKTGSGIDWSDEALTQO-----IKETIY 404

QY 404 VKLERMGL--TDLRCHIVTEYWTPLDIQAKYNSQSGIYGVVADRFRNKGFKAPQRSSEL 462

Db 405 RKLATIEVFEDIKSHVSEITIFENDEFQTYHAKFGSAFGMLPTLAQSNYYRPNVSRDY 464

QY 463 SNLYFVGGSVNPGGMPMVLTSQGLVRDKIVADLQ 497

Db 465 KDLYFAGASTHPGAGVPILVLTSAKITVDEMIDKIE 499

RESULT 12

US-10-369-493-18644

Sequence 18644, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 18644

LENGTH: 512

TYPE: PRT

ORGANISM: Halobacterium sp. NRC-1

US-10-369-493-18644

Query Match

Best Local Similarity 25.5%; Score 662.5; DB 15; Length 512;

Matches 169; Conservative 81; Mismatches 217; Indels 55; Gaps 11;

QY 9 VIVIGAGLGLSAAISLATAGFSVOLIEKNKDVGGKLNIMTKDGTFFDLGSPILTMPIH 68

Db 9 VAVIGAGGGSTACYLADAGADYVVEKTDQICGRASTLERDGRFDMGSPVLMEDVF 68

QY 69 BALFTGAGKNMADYVQIQVBEHWRNFPE-----DGSVIDLCEDAET 110

b 69 EBFASFDSHPSYGEITLDDPYRIFFKONEGRFRGRDAPGLNVDTDGDTTIDVTPDREQ 128  
y 111 QRRELDKLGSTYAQFORFLDYKSNLCTETECAGYFAKGLDGF-----WDLKFKYG 160  
b 129 VKQVFDAYEPGAGEVDDYLAQAKE-----NYVGEHFFVKTRPRVRDMDPKLAE 180  
y 161 PLRSLSFDVFRSMDQVRRFISDPKPLVILNFIKIVGSSPYDAPALMNLIPYIOYHYG 220  
c 181 YARGL---TLGSMQDHVEGYFDPHKLQVQMVTLVFLGSPDTTPALYNLMHSHVDFGLG 237  
y 221 LMYVKGWYCMQAKMEKLAVELGVEIELDAEVSIOKQGRACAVKAN--GDVLPADIV 278  
c 238 VYYPEGGMGVWDGITALARELGVEFRTHGHPVSAIK--GRRGGFKLDTPEGDAVLADV 294  
y 279 VNNEVIPAMEKLL---RSPASELKKMORFEPSCSEVLHLGVDRLYPQLAHNFFYSD 334  
c 295 VSDADYAHTEQALLPAKQKQYDADYDAKYAP--SAFLLYLGVGEVPELAHHTLVLP8 352  
y 335 HPREHFDVFKSHRLSDDPITYIVLVPCKTDPQAQAPAG-CEIKILPHIPLHDPDKLLFAE 393  
c 353 DWDTHPAQIFDDPAWDDPAVYLCVPSKTDVTAPDGHNSLFAVFPVAPGLD-DTPAVRE 411  
y 394 DYSALSERVLKLERMCLTDLRGHIYTEEVTPLDIOAKYVSNQSGSIYGVADRPNLGF 453  
c 412 SY---ADLVLDDEAENTGVDRDRIVEERFSVSEFADRYNSHQGTALGLAHTLRQTALP 468  
y 454 KAPQRSESLNLYFVGSGVNPGGEMVTLTGQLVDRKIVAD 495  
c 469 RPHRSTALDGLYFTGANTTGGIVGFMCLISGBVTANNVLAD 510

RESULT 13  
S-10-369-493-18983  
Sequence 18983, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Glodman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
CURRENT APPLICATION NUMBER: US/10/369,493  
PRIORITY FILING DATE: 2003-02-28  
PRIORITY FILING DATE: 2003-02-28  
PRIORITY FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 18983  
LENGTH: 499  
TYPE: PRT  
ORGANISM: Anabaena PCC7120  
S-10-369-493-18983

Query Match	25.2%	Score 654;	DB 15;	Length 499;
Best Local Similarity	32.2%;	Pred. No. 2.4e-56;		
Matches 158;	Conservative	112;	Mismatches 187;	Indels 34; Gaps 12
<i>f</i>	6	NQRIVIVIGAGLGLSRAISLATAGPVSOLLIRKNDKVGGLNMTXGDFPDLGPGSILTMP	65	
<i>s</i>	2	SKKVALVAGPGLATAIRLAGLYQVEIFEAAERVGGMRGPFVDSYAFDFTGPTILQIP	61	
<i>f</i>	66	HIPEALFTGAGKQMDYVOTQKVEPHRNFFEDSGSIDLCEAETQRRRLDKLGFQTYAQ	125	
<i>s</i>	62	HLVKELFEAGLNFADYVQKRLPYTRLKFWNDGTQDITSLQSPKQTALTRSDPLA	121	
<i>f</i>	126	FQRFPLDYKSNLCTETTAGTF-----AKGLDGFW-----DLKKYGPRLSLLSDFVSRWDQ	176	
<i>s</i>	122	FDRW--YSEHI-RKYSLGKPYLAGPARSIFGFLRPDELKMF-----LSFRPWNIIYQ	171	
<i>f</i>	177	GVRRFTSDPKLVEILNRYFKYVGSSPYDAPALMLLPYIOYHGLWYKGGHGYKQAQAME	236	

```

Db      172  HWMRFQDERLRYALSYSEKVLGVHPTVASSVVFSLIFLBFSGQVWHHPVGGFRALAQGLA 231
Qy      237  KLAVELGVRIERIDAVSIOQDGRACAVKLANDVLPAIVVYSNMKEVIPAMEKLLRSPA 296
Db      232  NAAQDLGVKHLHLSFVHQIWIQEQVGRGLELADASRQGFDTVVINADPAYAVRHLLPTSA 291
Qy      297  SEL---KMQREFPSCSLGVHLHGVDRLYPOLAHNFFYSDEHREHPDAVFKSHRLSD-D 352
Db      292  RGRYNTKLGQMGQFCSTFMLYLGNIRYEDLPHQIYLSLNTIRRLRPMWVDSALDETD 351
Qy      353  PTIYLVAPCKTPDPAQAPACGSIHKILPHIHPH---LDPDKLLTAEDYSALRERVLVKLER 408
Db      352  PPFVVCNPTIIDPSNAPAGHSTFLVLYPIENTSAYVDWD--IKQKSYT---DFILKRLHL 406
Qy      409  MGLTDLRQHIVTEWTFPLDLOAKYYSNQSGSIYGVADRPKNLG-PKAPORSSELSNLYF 467
Db      407  LGYHNIEQHIIVTQSCYTAQSWLLDYYRHLGAVFN-LSHNLTQLGPPPPPIRSENIAGLYW 465
Qy      468  VGGSVNPGGGM 478
Db      466  IGGAVHPGSEL 476

RESULT 14
US-10-369-493-20438
; Sequence 20438, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
; FILE REFERENCE: 38-19(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20438
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-20438

Query Match          24.8%; Score 645; DB 15; Length 498;
Best Local Similarity 30.8%; Pred. No. 1.9e-55;
Matches 151; Conservative 103; Mismatches 218; Indels 22; Gaps

```

Qy	10	IVITAGLGLSAAISLATAGFSVOLIENKQVGGKLNIMTKDQTFDLQPSILTMPIHE	69
Db	14	WVIGSFGCLAAAVRLGAKGYVTVLEKLGKAGRAYVHKQDQSFSDAQPTIVTADYLEE	73
Qy	70	ALFTGAGKNMADYVOIQKVEPHENFEDGSGVIDLCEDAETQRRLLDKLGRGYAOFQF	129
Db	74	ELWKLCGKMSDITLKPMSFFYIRFDGTHFDYSDDRDAVLQDTAKCFDDVPAYDRF	133
Qy	130	LDYSKNLCTETEGVYFAGK---LDGFMDLKFYGLRSLSSFDVFRSMQGVYRFRISDPK	186
Db	134	MAASHBI---FRVGFQEGDQFPFSHTDMLKI---APAMIKLESYRSYVGLVAKHFKDPK	187
Qy	187	LVEILNYFTKYVGSSPYDAPALMNLPIVOYHVGLMTVYKGMTCMAQMEKLAVELGVBI	246
Db	188	LQVPSFPHLLIGNPFWSSVYCLITYLEKQMGVHSMGTCALVTGLVNLIEGGNTI	247
Qy	247	RLDAAVESEIQKQDGBACAVKLANGDVLPADIVYSNMEVTPAMEKLRSPASELK----	301
Db	248	RYNQDVRQIVWENGTCACGVKLADGEVIKADIVSNADSASTYRYLL--SPETFRKWTDAK	305
Qy	302	MGRFPSPCSGLVHLHGVDRLYPQALHNFFYSORPHEHFDPAVFKSHRLSDDDPTIYLVAEP	361

Db 306 IEKSYNSLFWVYFGTERRVEDVKHHTILGPRKELISDIFSRKVAEDFSLYLRHT 365  
Qy 362 KTDPAQAGCEIILKILPHIL--DPKLLTADYSALSRVIVKLRMGLTDLRQHTV 419  
Db 366 ATDSLAPQGGDTTFVLSFVNLGLDTHWTKAETI---RASAKMLGATVLPDLNQIA 422  
Qy 420 TEETWPLDIQAKYNSQSTYGVVADRPKNLGFAPQRSSLSNLYFVGGSVNPGGMP 479  
Db 423 TSKITPTIDPQRLSSFGAAGLEPVLWQSAWFRPHNQSEDVAKRLYLVGAGTHPGAGLP 482  
Qy 480 MVTLSGQLVRDKIVAD 495  
Db 483 GV-ISSARVLDALVPE 497

RESULT 15  
US-09-547-267-5  
Sequence 5, Application US/09547267  
Patent No. US20020147371A1  
GENERAL INFORMATION:  
APPLICANT: Hohmann, Hans-Peter  
APPLICANT: Passamontes, Luis  
APPLICANT: Tessier, Michel  
APPLICANT: van Loon, Adolphus  
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: NJ  
COUNTRY: USA  
ZIP: 07110

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/547,267  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/660,645  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pokras, Bruce A.  
REGISTRATION NUMBER: 32,748  
REFERENCE/DOCKET NUMBER: RAN 6002/170  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-5801  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

Query Match 24.5%; Score 637; DB 9; Length 494;  
Best Local Similarity 31.2%; Pred. No. 1.2e-54;  
Matches 155; Conservative 90; Mismatches 233; Indels 18; Gaps 6;  
Qy 10 IVIGAGLGLGSAIATAGFSGVOLIEKNDKVGKLNIMTKDGFDFLGPISILTMPIHE 69  
Db 5 IVIGAGFGLALATLQAGIATTIVEARDXPGGAYVWNDQGHVFDAGTVVTDPSLR 64  
Qy 70 ALFTGAGKNMADYQIQKVEPHRNFFEDSGVIDLCEDAEYQRELDKLGPGTYAOFQR 129  
Db 65 ELWALSGQPMERDVTLLPVPFRLTWADGSRFEYVANDDDDELIRQVASFNPADVDGYRRF 124

Qy 130 LDYSKNLCETEAGYFAKGLDGFWDLLKFGYGLRSLILSFVFRSMDCQVRRFISDPKLV 189  
Db 125 HDYAEVYRE--GYLKLGTTPFLKLGMLNAAAPALMRLQAYRSVHSMVARFIQDPHLRQ 181  
Qy 190 ILNYFIKYVSSPYDAPALMMLLPYIOHYGLVYVKGMYGMAQAMEKLAVELGVEIRLD 249  
Db 182 APSFHTLLVGNPFTSTSIYALHALLERRGGVWFAKGTITQLVAGMVALFERLGGTLLN 241  
Qy 250 AEVSEIQKQDGRACAVKLANGDVLPAIVVSNMEVI PAMEKLL-----RSPASELKXMQ 303  
Db 242 ARVTRIDTEGDRATGVTLIDGRLRADTVASNGDVMSHYRDLILGHTRRGRTKAILNR-Q 300  
Qy 304 RPEPSCGLVHLGVDRILYPOLAHNFFYSQHPREHFDAYPKSHRLSDDPITVIVAPCKT 363  
Db 301 RW--SMSLFVLHFLGSLKRPENLAHSHVIFGPRYKGLVNEIFNGRPLDDDFSMYLSHSPCVT 358  
Qy 364 DPAQAPACGEIILKILPHIL---DPDKLLTADYSALSRVIVKLRMGLTDLRQHTV 420  
Db 359 DPLAPEGMSTHYVLAPVPHLGRADVDEAEAPGYA---ERIFEELERRAI PDLRKHLTV 415  
Qy 421 EETWPLDIQAKYNSQSTYGVVADRPKNLGFAPQRSSLSNLYFVGGSVNPGGMP 480  
Db 416 SRIETPADFSTELSAHGSAPSVPEPILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPG 475  
Qy 481 VTLGQLVRDKIVADL 496  
Db 476 VVGSAKATAQVMLSOL 491

Search completed: February 29, 2004, 15:27:52  
Job time : 34.998 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

in on: February 29, 2004, 14:35:44 ; Search time 15.1029 Seconds  
(without alignments)  
1698.885 Million cell updates/sec

File: US-09-941-947A-24

Effect score: 2598  
Sequence: 1 MNSNDNRVIVIGAGLGGLS.....MEMVTLSGQLVRDKIVADLQ 497

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Real number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*
- 1: /cgm2\_6/ptodata/2/iaa/5A COMB.pdp.\*
  - 2: /cgm2\_6/ptodata/2/iaa/5B COMB.pdp.\*
  - 3: /cgm2\_6/ptodata/2/iaa/6A COMB.pdp.\*
  - 4: /cgm2\_6/ptodata/2/iaa/6B COMB.pdp.\*
  - 5: /cgm2\_6/ptodata/2/iaa/ECTUS COMB.pdp.\*
  - 6: /cgm2\_6/ptodata/2/iaa/backfiles1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2598	100.0	497	4	US-09-934-903-18
2	703.5	27.1	489	1	US-08-095-726-10
3	703.5	27.1	489	1	US-08-096-043-10
4	703.5	27.1	489	1	US-08-096-623A-10
5	702.5	27.0	489	1	US-08-095-726-8
6	702.5	27.0	489	1	US-08-095-726-8
7	702.5	27.0	489	1	US-08-096-623A-8
8	671	25.8	492	1	US-07-783-705A-4
9	637	24.5	494	3	US-08-660-645A-5
10	637	24.5	494	3	US-09-298-718-5
11	637	24.5	494	3	US-09-546-969-5
12	637	24.5	494	3	US-08-980-832-4
13	637	24.5	494	4	US-09-547-267-5
14	637	24.5	494	4	US-09-920-923B-4
15	618.5	23.8	582	4	US-09-091-725-17
16	607.5	23.4	511	4	US-09-934-903-16
17	596	22.9	290	4	US-08-936-165A-462
18	184.5	7.1	610	4	US-09-443-184-55
19	155.5	6.0	508	1	US-08-472-028A-4
20	155.5	6.0	508	2	US-08-808-931-4
21	155.5	6.0	508	3	US-08-808-323-4
22	155.5	6.0	508	3	US-09-050-603A-4
23	155.5	6.0	508	3	US-09-102-420B-4
24	155.5	6.0	508	3	US-09-071-296-4
25	155.5	6.0	508	3	US-09-196-268-4
26	155.5	6.0	508	3	US-09-015-683-4
27	155.5	6.0	508	4	US-09-191-998-4

28 155.5 6.0 508 4 US-09-497-698-4 Sequence 4, Appli  
29 150.5 5.8 598 3 US-09-352-159-42 Sequence 42, Appl  
30 149.5 5.8 578 4 US-09-690-942-10 Sequence 10, Appl  
31 144 5.5 524 3 US-09-230-388-1 Sequence 1, Appli  
32 144 5.5 524 4 US-09-912-176-1 Sequence 1, Appli  
33 143.5 5.5 571 4 US-09-690-942-15 Sequence 15, Appl  
34 143 5.5 598 3 US-09-352-159-40 Sequence 40, Appl  
35 143 5.5 991 3 US-09-352-159-27 Sequence 27, Appl  
36 143 5.5 991 3 US-09-352-168-27 Sequence 27, Appl  
37 143 5.5 1000 3 US-09-352-159-25 Sequence 25, Appl  
38 143 5.5 1000 3 US-09-352-168-25 Sequence 25, Appl  
39 143 5.5 1196 3 US-09-352-159-31 Sequence 31, Appl  
40 143 5.5 1196 3 US-09-352-168-31 Sequence 31, Appl  
41 143 5.5 1205 3 US-09-352-159-29 Sequence 29, Appl  
42 143 5.5 1205 3 US-09-352-168-29 Sequence 29, Appl  
43 141.5 5.4 462 3 US-09-352-159-8 Sequence 8, Appli  
44 141.5 5.4 462 3 US-09-352-168-8 Sequence 8, Appli  
45 141.5 5.4 829 3 US-09-352-159-33 Sequence 33, Appl

#### ALIGNMENTS

RESULT 1  
US-09-934-903-18  
; Sequence 18, Application US/09934903  
; Patent No. 6660507  
; GENERAL INFORMATION:  
; APPLICANT: Koffas, Mattheos  
; APPLICANT: Odom, J. Martin  
; APPLICANT: Schenzle, Andreas J.  
; APPLICANT: No. 6660507ton, Kelley C.  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Rovierto, Pierre  
; APPLICANT: Picataggio, Stephen  
; APPLICANT: Cheng, Qiong  
; TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production  
; FILE REFERENCE: CL1646 US NA  
; CURRENT APPLICATION NUMBER: US/09/934, 903  
; CURRENT FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/229, 907  
; PRIOR FILING DATE: September 1, 2001  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 18  
; LENGTH: 497  
; TYPE: PRT  
; ORGANISM: Methylobionas 16a  
; FEATURE:  
; OTHER INFORMATION: Amino acid sequences encoded by ORF9  
US-09-934-903-18

Query Match 100.0%; Score 2598; DB 4; Length 497;  
Best Local Similarity 100.0%; Pred. No. 8.8e-259;  
Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSNDNRVIVIGAGLGGLSAAISLATAGSFVQLIEKNDKVGKLNIMTKDGFDFDLGSP 60  
DB 1 MNSNDNRVIVIGAGLGGLSAAISLATAGSFVQLIEKNDKVGKLNIMTKDGFDFDLGSP 60  
QY 61 ILTMTHI FEALPTGAGKNADYVQIQKVEPHWRNPFDFGSDVIDLCEDAETORRELDKLG 120  
DB 61 ILTMTHI FEALPTGAGKNADYVQIQKVEPHWRNPFDFGSDVIDLCEDAETORRELDKLG 120  
QY 121 GTTACQFORFLDYSKNLTCTETBAGTFYAGKLDGFWLLKPYGFLRSLLSDFVFRSMDDQVRR 180  
DB 121 GTTACQFORFLDYSKNLTCTETBAGTFYAGKLDGFWLLKPYGFLRSLLSDFVFRSMDDQVRR 180  
QY 181 FSDPKLVEILNYFKYVGSPPDAPALNNLLPYQHYGLWYVKGMYGMAQAMEKLAV 240  
DB 181 FSDPKLVEILNYFKYVGSPPDAPALNNLLPYQHYGLWYVKGMYGMAQAMEKLAV 240  
QY 241 ELGVEIRLDAEYSEIQKQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLLRSPASELK 300

db 241 ELGVIRKDAEVESEIQKQGRACAVKANGDVLPAIVVSNMEVIPAMEKLRSPASELK 300  
iy 301 KMORFEPSCGLVHLGVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVA 360  
b 301 KMORFEPSCGLVHLGVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVA 360  
y 361 CKTDPAQAPAGCEIILKILPHILDPKLLTAEDYSALRERVLVLERMGLTDLRQHI 420  
b 361 CKTDPAQAPAGCEIILKILPHILDPKLLTAEDYSALRERVLVLERMGLTDLRQHI 420  
y 421 EEWTPLDIOAKYYSNQGSIYGVVADRFKNLGPAPQRSSELSNLYFVGGSYNPGGMP 480  
b 421 EEWTPLDIOAKYYSNQGSIYGVVADRFKNLGPAPQRSSELSNLYFVGGSYNPGGMP 480  
y 481 VTLSGQLVRDKIVADLQ 497  
b 481 VTLSGQLVRDKIVADLQ 497

## RESULT 2

S-08-095-726-10

Sequence 10, Application US/08095726

Patent No. 5530189

## GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L  
APPLICANT: Brinkhaus, Friedhelm L  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H  
APPLICANT: Yarger, James G  
APPLICANT: Yen, Hwei-Che B

TITLE OF INVENTION: Beta-Carotene Biosynthesis in  
Genetically Engineered Hosts

NUMBER OF SEQUENCES: 79

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
STREET: 200 E Randolph St

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60680-0703

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/095,726

FILING DATE: 21-JUL-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/785,566

FILING DATE: 30-OCT-1991

ATTORNEY/AGENT INFORMATION:

NAME: Galloway, No. 5530189val B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 3128567180

TELEFAX: 3128564572

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 489 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-095-726-10

Query Match 27.1%, Score 703.5; DB 1; Length 489;

Best Local Similarity 32.9%; Pred. No. 1.5e-63;

Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

2y 7 ORVTVIGAGLGLSAISLATAGSVOLIEKNDKVGKLMIMTKQGTFTDLGPSILTMPH 66

db 2 EKTVVIGAGLGLAIALQLQAGIPTVLLQRDKPGGRAYVWHDQGTFTDAGFTVITDPT 61

Qy 67 IFPALFTGAGKQNMADYVOIQKVEPHRNFEEDSGVIDLCEDAETQRELDKLGPGTYAOF 126  
Db 62 ALFALFTLAGRMEDVRLLPVKPFTYRLCWESGKTLDYANDSFLELEAQITQNPDRVEGI 121  
Qy 127 QRFLDYSKNLCTEAGYFAKG---LDGFWDLLIKFYGPLRLSLSPDFVFRSMDQGVRRFIS 183  
Db 122 RRFAYSQAVFQE---GYLRIGSVPPFLSFRDMLR-AGP---QLLKLQAMQSVYQSVSRFIE 175  
Qy 184 DPKLVEILANFYKYVSSPYDAPALNNLLPYQYHGLVYVKGWYGMQAQMEKLAVELG 243  
Db 176 DEHLRQAFPSHLLVGNFTTSSIIYTLHALEREWGVFPFGGTGALVNGWVLFLLFDLG 235  
Qy 244 VEIRLDAEVESEIQKQGRACAVKANGDVLPAIVVSNMEVIPAMEKLRSPASELK-- 300  
Db 236 GEIELMARVEELVADNRVSVQVELADGRIFDITDAVASNADVNTYKLLGTIPVGOKRAA 295  
Qy 301 KMORFEPSCGLVHLGVDRLYPOLAHNFFYSDHPREHFDVAFKSHRLSDDDPTIYLVA 360  
Db 296 RLBRKSMNSLFLVYFLNQLPHSLFIIEPPSLTQGLFA-----ANATRHNSNLYLVAAGT 355  
Qy 361 CKTDPAQAPAGCEIILKILPHILDPKLLTAEDYSALRERVLVLERMGLTDLRQHI 420  
Db 356 CVTDPSLAPPCCASFYVLAPVPHLGNAPLDWQEGPKLRDRIFDYLEERYMPLGRSLQVLT 415  
Qy 421 EEWTPLDIOAKYYSNQGSIY-----GVVADRFKNLGPAPQRSSELSNLYFVGGSV 472  
Db 416 QRIFTRQTSRHWAILGSLFIEPPSLTQGLFA-----ANATRHNSNLYLVAAGT 464  
Qy 473 NPGGMPMTLSGQLVRDKIVADLQ 497  
Db 465 HPGAGIPGVVGLAEASTASLMIEDLQ 489

RESULT 3  
US-08-096-043-10  
Sequence 10, Application US/08096043  
Patent No. 5530189  
GENERAL INFORMATION:  
APPLICANT: Ausich, Rodney L  
APPLICANT: Brinkhaus, Friedhelm L  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H  
APPLICANT: Yarger, James G  
APPLICANT: Yen, Hwei-Che B  
TITLE OF INVENTION: Lycopen Biosynthesis in  
Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
STREET: 200 E Randolph St  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60680-0703  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,043  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/785,568  
FILING DATE: 30-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, No. 5530189val B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 3128567180  
TELEFAX: 3128564572  
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 489 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 3-08-096-043-10

Query Match 27.1%; Score 703.5; DB 1; Length 489;  
 Best Local Similarity 32.9%; Pred. No. 1.5e-63;

Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

7 QKVIVIGAGLGLSLNAISLATAGSVQLIKNDKVGKLNIMTKDGTDFLGPISILTMHP 66  
 :::::::::::::::::::::  
 2 EKTWVIGAGFGLALAIRLOAGIPTVLLEQDRPGGRAYVWHDDQGFDFAGTVIDPT 61  
 :::::::::::::::::::::  
 67 IFEALFTGAGKNADYVOIQVEPHWRNFFDGSVIDLCEDAETQRELDKLGPGTVAQF 126  
 :::::::::::::::::::::  
 62 ALAULTLAGRMEDYVRLPVKFFYLCSWGSKTLDYANDSFELEAQITQFNPRDVEGY 121  
 :::::::::::::::::::::  
 127 QRLDYSKNLCTETAGYFAKG---LDGFMDLLKFPGLRSLLSFDVFRSMDQGVRRFIS 183  
 :::::::::::::::::::::  
 122 RFLAYSQAVFOE---GYLRGSPVFLSFRDMLR-AGP---QLLKLAQMSVYQSVSRFIE 175  
 :::::::::::::::::::::  
 184 DPKLVEILNYFKYVSSPYDAPALMNLPIQYHGLVYVKGGMGMAQAMEKLAVELG 243  
 :::::::::::::::::::::  
 176 DEHLRQAFSFLSLVGGNPTTSSYITLIHALEREWGVPFEGGTGALVNGVKLFTDLG 235  
 :::::::::::::::::::::  
 244 VEIRDAEVSEIQKODGRACAVKLANGDVLPAIVWSNMEVIPAMEKLLRS-PASELK-- 300  
 :::::::::::::::::::::  
 236 GEIELNARVEELVWADNRVSOVLADGRIFDTDAVASNADVVNTYKLLGTIPVGQKRAA 295  
 :::::::::::::::::::::  
 301 KMQRPEPCSGVLVHLGVDRLYPOLAHNFFYSDHPREHDAVFKSHRLSDPTIYLAP 360  
 :::::::::::::::::::::  
 296 RLERKSMNSLFLVYFGNLQPHSLAHHTICGPRYRELIDEIFTGSALADDFSLYLHSP 355  
 :::::::::::::::::::::  
 361 CKTDPAPAGACEIILKILPHILDPDKLTAEDYSALRERVLVKLRMGITDLRQHIVT 420  
 :::::::::::::::::::::  
 356 CVTDFSLAPPCASFYVLAPVPHLGNAPLWAQSGPKLRDRIFDYLEERYWGLRSQVLT 415  
 :::::::::::::::::::::  
 421 EBYWTPLOIQAKYYSNQSIIY-----GVVADRPNLGFKAQORSSELSNLYFVGGSV 472  
 :::::::::::::::::::::  
 416 QRIFTQTSRHAWIAILGSLFIEPSPITQGLFA-----ANATRSNLYLVAAGT 464  
 :::::::::::::::::::::  
 473 NFGGGMWVTLSSGQLVRDKIVADLQ 497  
 :::::::::::::::::::::  
 465 HPGAGIPGVVGLAESTASIMIEDLQ 489  
 :::::::::::::::::::::

35ULT 4

3-08-096-623A-10

Sequence 10, Application US/08096623A

Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Asich, Rodney L.

APPLICANT: Brinkhaus, Friedhelm L.

APPLICANT: Mukharji, Indrani

APPLICANT: Proffitt, John H.

APPLICANT: Yarger, James G.

APPLICANT: Yen, Huel-Che B.

TITLE OF INVENTION: Biosynthesis of Zeaxanthin and

TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts

NUMBER OF SEQUENCES: 104

CORRESPONDENCE ADDRESS:

ADDRESSER: Welsh & Katz, Ltd.

STREET: 120 S. Riverside Plaza, 22nd Floor

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/096.623A

FILING DATE: 22-JUL-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/805,061

FILING DATE: 09-DEC-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/662,921

FILING DATE: 28-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/562,674

FILING DATE: 03-AUG-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/525,551

FILING DATE: 18-MAY-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/487,613

FILING DATE: 02-MAR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Gamson, Edward P.

REGISTRATION NUMBER: 29,381

REFERENCE/DOCKET NUMBER: AMO-006.1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 655-1500

TELEFAX: (312) 655-1501

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 489 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-096-623A-10

Query Match 27.1%; Score 703.5; DB 1; Length 489;

Best Local Similarity 32.9%; Pred. No. 1.5e-63;

Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

QY 7 QKVIVIGAGLGLSLNAISLATAGSVQLIKNDKVGKLNIMTKDGTDFLGPISILTMHP 66  
 DB 2 EKTWVIGAGFGLALAIRLOAGIPTVLLEQDRPGGRAYVWHDDQGFDFAGTVIDPT 61  
 QY 67 IFEALFTGAGKNADYVOIQVEPHWRNFFDGSVIDLCEDAETQRELDKLGPGTVAQF 126  
 DB 62 ALAULTLAGRMEDYVRLPVKFFYLCSWGSKTLDYANDSFELEAQITQFNPRDVEGY 121  
 QY 127 QRLDYSKNLCTETAGYFAKG---LDGFMDLLKFPGLRSLLSFDVFRSMDQGVRRFIS 183  
 DB 122 RFLAYSQAVFOE---GYLRGSPVFLSFRDMLR-AGP---QLLKLAQMSVYQSVSRFIE 175  
 QY 184 DPKLVEILNYFKYVSSPYDAPALMNLPIQYHGLVYVKGGMGMAQAMEKLAVELG 243  
 DB 176 DEHLRQAFSFLSLVGGNPTTSSYITLIHALEREWGVPFEGGTGALVNGVKLFTDLG 235  
 QY 244 VEIRDAEVSEIQKODGRACAVKLANGDVLPAIVWSNMEVIPAMEKLLRS-PASELK-- 300  
 DB 236 GEIELNARVEELVWADNRVSOVLADGRIFDTDAVASNADVVNTYKLLGTIPVGQKRAA 295  
 QY 301 KMQRPEPCSGVLVHLGVDRLYPOLAHNFFYSDHPREHDAVFKSHRLSDPTIYLAP 360  
 DB 296 RLERKSMNSLFLVYFGNLQPHSLAHHTICGPRYRELIDEIFTGSALADDFSLYLHSP 355  
 QY 361 CKTDPAPAGACEIILKILPHILDPDKLTAEDYSALRERVLVKLRMGITDLRQHIVT 420  
 DB 356 CVTDFSLAPPCASFYVLAPVPHLGNAPLWAQSGPKLRDRIFDYLEERYWGLRSQVLT 415  
 QY 421 EBYWTPLOIQAKYYSNQSIIY-----GVVADRPNLGFKAQORSSELSNLYFVGGSV 472  
 DB 416 QRIFTQTSRHAWIAILGSLFIEPSPITQGLFA-----ANATRSNLYLVAAGT 464  
 QY 473 NFGGGMWVTLSSGQLVRDKIVADLQ 497

465 HPGAGIPGVVGLAESTASLMIEDLQ 489

RESULT 5

US-08-095-726-8  
Sequence 8, Application US/08095726  
Patent No. 5530188

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L  
APPLICANT: Brinkhaus, Friedhelm L  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H  
APPLICANT: Yarger, James G  
APPLICANT: Yen, Hwei-Che B  
TITLE OF INVENTION: Beta-Carotene Biosynthesis in  
TITLE OF INVENTION: Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
STREET: 200 E Randolph St  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/095,726  
FILING DATE: 21-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/785,566  
FILING DATE: 30-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, No. 5530188val B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 3128567180  
TELEFAX: 3128564972  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 489 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-095-726-8  
Query Match 27.0%; Score 702.5; DB 1; Length 489;  
Best Local Similarity 32.9%; Pred. No. 1.9e-63;  
Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

7 QRVIVIGAGLGGLSAAISLATAGFSVQLIEKNDKVGKLNIMTKDFTFDLQPSILTMHP 66  
2 KKTVVIAGFGGLALALRLQAGIPTVLLQQRDKPGRAVYVHWDQGTFDAGPTVIIDPT 61  
67 IFEALFTGAGNMADYVQIQKVEPHNRNFFEDSGSVIDLCEADTORRELDKLGPGTYAQF 126  
62 ALEALFTLAGRMEDYVRLLPVKPFYRLCWESGKTLDYANDSFLEAQITQPNPRDVGY 121  
127 ORFLDYSKNLCTEAGYFAGK--LDGFWDLLKFGYPLRSLLSPVFRSMQGVRRFIS 183  
122 RRLFLAYCAVQOE--GYRLGVSVPFLSFEDMLR-AGP--QLLKLAQWSVYQSVSRFIE 175  
184 DPKLVEILNYFIKTVGSSPDAPALMNLPIYIQHVLGWLTVYKGYMGMAQAEKLAVELG 243  
176 DEHLRQAFSPHLLVGNPFTTSIYTLIALEREWGVPPEGGTALYNGMVKLFTDLG 235  
244 VEIRLDRAVSIQKQDGRACAVKLANGDVLPAITVSNMMEVIPAMEKILLES-PASELK-- 300  
236 GEIELNARVEELVADNRVSRVLADGRIFDITDAVASNADVNVTYKLLGTIPVQKRAA 295

301 KMORFSPSCSLVHLGVDRLYQLAHNFFYSDEHREHFDVFKSHRLSDDDPTIYLVAP 360  
296 RLEKKSMSNSLFLVYFGNQPHSQLAHHTICGPRYRELIDEIFTGSAADDFSLYHSP 355  
361 CKTDPAAQAPACBIIKILPHIPHLDPKLLTADVSALRERVVLKLERMGLTDLRQHIYT 420  
356 CVTDPSLAPPCCASFVYLAFLAPVPHLGNAPLDWAQEGPKLRDRIFDVLEERYMPLRSQLYT 415  
421 BEYWTPLDIAQKYNSQGSY-----GVVADRFKNLGFKAQORSSLSNLYFYGGSV 472  
416 QRIFRTQTSRHAWIALIGSLFIEPPSLTQGLFA-----ANATHSNLYLVAAGT 464  
473 NPGGEMVMTLSQGLVRDKIVADLQ 497  
465 HPGAGIPGVVGLAESTASLMIEDLQ 489

RESULT 6

US-08-096-043-8  
Sequence 8, Application US/08096043  
Patent No. 5530189

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L  
APPLICANT: Brinkhaus, Friedhelm L  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H  
APPLICANT: Yarger, James G  
APPLICANT: Yen, Hwei-Che B  
TITLE OF INVENTION: Lycopene Biosynthesis in  
TITLE OF INVENTION: Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
STREET: 200 E Randolph St  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,043  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/785,568  
FILING DATE: 30-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, No. 5530189val B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 3128567180  
TELEFAX: 3128564972  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 489 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-096-043-8  
Query Match 27.0%; Score 702.5; DB 1; Length 489;  
Best Local Similarity 32.9%; Pred. No. 1.9e-63;  
Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

7 QRVIVIGAGLGGLSAAISLATAGFSVQLIEKNDKVGKLNIMTKDFTFDLQPSILTMHP 66  
2 KKTVVIAGFGGLALALRLQAGIPTVLLQQRDKPGRAVYVHWDQGTFDAGPTVIIDPT 61  
67 IFEALFTGAGNMADYVQIQKVEPHNRNFFEDSGSVIDLCEADTORRELDKLGPGTYAQF 126



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

on on: February 29, 2004, 14:33:49 ; Search time 43.7446 Seconds  
(without alignments)  
3837.172 Million cell updates/sec

tie: US-09-941-947a-38

fect score: 2768

quence: 1 MSFIDAVVGGSHALVSA.....ALSDSRGKASQMMRRSSRS 532

oring table: BL0SUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

SPTREMBL 25:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_virus:  
16: sp\_bacteriap:  
17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	859.5	31.1	511	16	Q9RY57
2	688	24.9	542	16	Q55808
3	660.5	23.9	565	16	Q8YQ55
4	624.5	22.6	581	4	Q8YRQ1
5	624.5	22.6	581	4	Q8N2H3
6	572.5	20.7	556	10	Q8YVK3
7	562.5	20.3	647	10	Q9FGZ1
8	557.5	20.1	521	16	Q98FP6
9	538.5	19.5	543	16	Q9A3T1
10	537.5	19.4	519	16	Q98FP7
11	533	19.3	518	17	Q97W24
12	525	19.0	517	17	Q974W1
13	509	18.4	539	16	Q8FP01
14	497	18.0	545	16	Q89FW3
15	469	16.9	524	16	Q98BS8
16	466.5	16.9	544	5	O61196

17 443 16.0 544 16 Q82PC2 streptomyc  
18 441.5 16.0 539 16 Q9L187 streptomyc  
19 439.5 15.9 538 17 Q9YCC0 aeropyrum p  
20 418 15.1 489 16 Q9RYF5 deिनococcus  
21 385.5 13.9 540 16 Q9K414 streptomyc  
22 363.5 13.1 523 16 Q98BS6 rhizobium 1  
23 359 13.0 472 2 Q9EY15 streptomyc  
24 349.5 12.6 482 16 Q8FSA2 corynebacte  
25 348 12.6 489 2 Q9KHE1 streptomyc  
26 340 12.3 472 16 Q9ZBG4 streptomyc  
27 336.5 12.2 535 16 Q8XIC4 ralstonia s  
28 335 12.1 472 16 Q82LS4 streptomyc  
29 318.5 11.5 536 16 Q7TVK4 mycobacteri  
30 314.5 11.4 536 16 Q07794 mycobacteri  
31 309 11.2 533 16 Q82NN7 streptomyc  
32 302 10.9 499 2 Q9K566 mycobacteri  
33 296.5 10.7 471 16 Q8NT31 corynebacte  
34 293.5 10.6 471 16 Q82NF9 streptomyc  
35 284 10.3 473 16 Q06826 mycobacteri  
36 284 10.3 473 16 Q7U032 mycobacteri  
37 283 10.2 480 16 Q53244 mycobacteri  
38 276.5 10.0 588 2 Q83X01 streptomyc  
39 275 9.9 523 16 Q9RIX9 streptomyc  
40 274 9.9 507 2 Q9EXL0 streptomyc  
41 273 9.9 480 16 Q7TXH1 mycobacteri  
42 270 9.8 501 16 Q55455 synchocyst  
43 264.5 9.6 517 2 Q9ZGE4 heliobacill  
44 262.5 9.5 520 16 Q7V6U2 streptomyc  
45 261 9.4 513 16 Q93HP2 streptomyc

#### ALIGNMENTS

#### RESULT 1

Q9RY57 PRELIMINARY; PRT; 511 AA.  
ID Q9RY57  
AC Q9RY57;  
DT 01-MAY-2000 (TREMUREL. 13, Created)  
DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)  
DT 01-JUN-2003 (TREMUREL. 24, Last annotation update)  
DE Phytoene dehydrogenase, putative.  
GN DR0093.  
OS Deinococcus radiodurans.  
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
OC Deinococcaceae; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1", 1577 (1999).  
RL Science 286:1571-1577 (1999).  
DR EMBL; AE001872; AAF09686.1; -.  
DR PIR; E75561; E75561.  
DR TIGR; DR0093; -.  
DR InterPro; IPR008151; Phyt\_n\_dehydro.  
DR ProDom; PD139017; Phyt\_n\_dehydro; 1.  
KW Complete proteome.  
SQ SEQUENCE 511 AA; 56527 MW; 7DCC3FB1D79BEB9CD CRC64;

Query Match 31.1%; Score 859.5; DB 16; Length 511;  
Best Local Similarity 39.7%; Pred. No. 1.9e-45;  
Matches 207; Conservative 80; Mismatches 208; Indels 27; Gaps 11;







```

367 -----LPPRAVLGMSF-SGIDPTIAPAGRHQVTLMSQMPYRLS-GHRDWSVAEAE 417
395 CHDSNGLPSSRRPVIEITPSTLNTISPPCKHVINLFIQYTFKPSDGSWEDPTREAF 454
418 ADRIVGENEAFAPGFTDVLDRFTQTPRDISELSGMTGGNMVHVMESLDMQMLWRPLPEL 477
455 AQRCPKLDIYAPGSSIIISYDMLTPDLEREIGLTGGNIFHGAMGLDSLFLMRPVKGM 514
478 SCHRVPGADGLYLTCASHTPGGVSAGSRAARIALSDSR 519
515 SNYRSP-LKGLYLCGSGHFGGVMGAPGRNAHVLDLKR 555

RESULT 7
PGZ1
Q9FGZ1 PRELIMINARY; PRT; 647 AA.
01-MAR-2001 (TRENBLrel. 16, Created)
01-MAR-2001 (TRENBLrel. 16, Last sequence update)
01-OCT-2003 (TRENBLrel. 25, Last annotation update)
Phytoene dehydrogenase-like.
Arabidopsis thaliana (Mouse-ear cress).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eurosids II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
STRAIN=Columbia;
MEDLINE=20181125; PubMed=10718197;
Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
Tabata S.,
Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
features of the regions of 3,076,755 bp covered by sixty P1 and TAC
clones."
DNA Res. 7:31-63(2000).
EMBL; AB023033; BAB10768.1; -.
InterPro; IPR000205; NAD_B5.
InterPro; IPR008151; Phytin dehydro.
ProDom; PD139017; Phytin dehydro; 1.
SEQUENCE 647 AA; 70687 MW; 3DC4C5918A8D0FC6 CRC64;

Query Match 20.3%; Score 562.5; DB 10; Length 647;
Best Local Similarity 30.9%; Pred. No. 7.9e-27;
Matches 179; Conservative 78; Mismatches 213; Indels 109; Gaps 20;

6 DAVVGGSHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERFPYKVDGSSAHMTR 65
15 DAVVGGSHNGLTAAAYLARGGLSVAVLERRHVIGGAATVEETVPGFKFSRCSYLOGLLR 74
66 HSGIIEELGLGAHLRYIDCDPWAFAPAPGTDGPIVPHRDLDATCQSIERACGFKDAD 125
75 PC-IIRLELGRHGLKLLKSPSP---TCLDGRVLLGLPOQDLNHSRISK-FSKHDAD 129
126 AYRFVAVWSERSRHYMKAF-----STPTGSLNLIAGFGLATARGNSLSRQ----- 173
130 AYPYR-----EKOLRFGCFWMDPLDSTTPESLQSASSF-----NDKLSNMYKSAF 176
174 -----FLAPCDALLDEYFDESEALKAAALNFGA-QSGPPMSRPG 210
177 WARCIRQAVSLGHKDMVAFMDLLILAFASKVINNNFESVDVILKASLATDAVIGSTASVHTPG 236
211 TAPMVGFAALMHVLP-----GRAVGGGALSAAALSRMAVDGATVALGGVTSIRR 262
237 S-----GVVLLHHVWGETGKEGWSYVEGGMGSVSMALANAKEAGAEFTNAEYSEI-- 290
263 NSNFWYTTESGREVHARKVIAGCHILTTLLDLCNGGFORTLLDHRKIRVPGPGIGAV- 321
291 -----LTDSSSTVKGVLADGTRVRESSA-ILSNATPYRTYVELVPTNVLPENFVSAIK 342
322 -----LRLATSPALSVRGDATTRESTSGLLQLLVSDRAHLRTAH-GAALAGE---- 366
343 NSDYSSATTINLAVDKLPQFQ-CCNINSHGCPGE-----HFGTIHIGASSMDVEHSA 394

```

```

QY 367 -----LPPRAVLGMSF-SGIDPTIAPAGRHQVTLMSQMPYRLS-GHRDWSVAEAE 417
DB 395 CHDSNGLPSSRRPVIEITPSTLNTISPPCKHVINLFIQYTFKPSDGSWEDPTREAF 454
QY 418 ADRIVGENEAFAPGFTDVLDRFTQTPRDISELSGMTGGNMVHVMESLDMQMLWRPLPEL 477
DB 455 AQRCPKLDIYAPGSSIIISYDMLTPDLEREIGLTGGNIFHGAMGLDSLFLMRPVKGM 514
QY 478 SCHRVPGADGLYLTCASHTPGGVSAGSRAARIALSD 516
DB 515 SNYRSP-LKGLYLCGSGHFGGVMGAPGRNAHVLTNN 552

RESULT 8
Q98FP6 PRELIMINARY; PRT; 521 AA.
AC Q98FP6;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DB Phytoene dehydrogenase.
GN MLR3676
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAPEF303099;
RA MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
Takeuchi C., Yamada M., Tabata S.;
"Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti";
DNA Res. 7:331-338(2000).
EMBL; AP003002; BAB50521.1; -.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR002937; Amino oxidase.
DR InterPro; IPR000205; NAD_B5.
DR InterPro; IPR008151; Phytin dehydro.
DR InterPro; IPR003042; Rng_aminoxigenase.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00420; RINGMINOXGNASE.
DR ProDom; PD139017; Phytin dehydro; 1.
KW Complete proteome.
SQ SEQUENCE 521 AA; 54661 MW; 12846C9155B8197 CRC64;

Query Match 20.1%; Score 557.5; DB 16; Length 521;
Best Local Similarity 34.2%; Pred. No. 1.2e-26;
Matches 188; Conservative 75; Mismatches 213; Indels 73; Gaps 28;

QY 1 MSATLDAVVGSHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERFPYKVDGSSA 60
DB 1 MTSP-DAVIGGGHNGVAAATLAKAGKVLVLEASVEGGAARTBEFAPGFRV--SSIA 57
QY 61 HLMTR-HSGIIEELGLGAHLRYIDCDPWAFAPAP-GTDPGIVPHRDLDATCQSIERA 118
DB 58 HLLNRLHPDVVKTELETHGLQPARAD---FLPSAALSXKDGAPALMLH-----GA 103
QY 119 CGTKDADAYRRFVAVNSE-----RSHVMKAPST--PP--TGSNLI--GAFG--GLATA 164
DB 104 YGEVLTGASFSSEGSAMKDLRAQLLRVAGILKPLFTRPPDLAGSLMETASLGQTALAK 163
QY 165 RGNSELSRQFLA-----PGDALLDEYFDESEALKAAALAW---FGAQSGP--PMSEPG---- 210
DB 164 KLGGEDMRDFLRVLNMXVAD-LLDEQLRDRDLKGLLAFOATLGLSHLGPSPSTLLGLYYR 222
QY 211 TAPMVGFAALMHVLPGRVGGGALSAAALSRMAVDGATVALGGVTSIRRSNWT-V 269

```

```

b 223 LAGEAGGAAGVLPQ---GCGAGVAAIRAAAEKAGVTIRTSVAAAKIIVEKGHAVGV 278
y 270 TTESGREHARVIAAGCHILTT-DLLG----NGGFDRTTLDHWRKIRVGGIGAVLRL 324
b 279 TLGTGELRARTVVSAINPATTFLDVGPREIDTGFVRKV-----KTRM-KGDAAKLHL 332
y 325 ATSAIPSYRG-DATTRESTSGQLIVSDRAHLRTAAGALAGELPPRPVAVLCMGFSGI-D 382
c 333 ALDRSPQFTGADAAAHK---GRLVAPDPDHVERAFNECKYGEFSPD-VLEITLPSLAD 388
y 383 PTIAPAGRGQVTLWSQWQPYRLSGHRDWSVAEABADRIVGEMEAFAFGDTSVLDRFIQ 442
b 389 PSLAPDGCVL SAVVOYAFYALK--EGWAAGKPKFIKAVMAQLEAYAGICKSVVHAE LL 446
y 443 TPRDISESGYIGGNVWHVEMSLDQMLWRPLPELSCHRVPGADGLYLTGASTHPGGVS 502
b 447 TPADIETRYNPGGHHGELQDQMLSRPVGWSGYDTP-VEGLFLAGAGSHPGGGSVS 505
y 503 GASGRSAAR 511
c 506 GAFGLNAAR 514

RESULT 9
9A3T1
c Q9A3T1 PRELIMINARY; PRT; 543 AA.
t 01-JUN-2001 (TrEMBLrel. 17, Created)
t 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
t 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
e Phytoene dehydrogenase-related protein.
n NCBI_TaxID=155892;
s Caulobacter crescentus.
c Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
c Caulobacteraceae; Caulobacter.
x NCBI_TaxID=155892;
n [1]
p SEQUENCE FROM N.A.
c STRAIN=ATCC 19089 / CB15;
c MEDLINE=21173698; PubMed=11259647;
a Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
a Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
a Potocka I., Nelson W.C., Newton A.C., Stephens C., Phadke N.D., Ely B.,
a DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
a Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
a Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
a Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
t "Complete genome sequence of Caulobacter crescentus.";
l Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
r EMBL; AB005975; AAK25083.1; -.
r PIR; G87635; G87635.
r TIGR; CC3121; -.
r InterPro; IPR008151; Phyt_n_dehydro.
r ProDom; PD139017; Phyt_n_dehydro; 1.
w Complete proteome.
q SEQUENCE 543 AA; 57881 MW; 3FE9EB19403B1DE CRC64;

Query Match 19.5%; Score 538.5; DB 16; Length 543;
Best Local Similarity 33.1%; Pred. No. 1.9e-25;
Matches 180; Conservative 76; Mismatches 234; Indels 53; Gaps 21;

y 6 DAVVCSGHNALVSAYLAREGWSVEVLEKDTVLGAVSTVERFPYKVDKGSSAHLMI 65
b 8 DAVIIGGHNGLVCAYLAKAGLKVTVCARGVVGGAATVEFFHGFGR-NSVASTVSL 66
y 66 HSGIIBELGAGHGLRYIDCDPWAFAPAGTDPGPIVFHRLDTCQSIERACTKQAD 125
b 67 NRPVADMGRLGELGFLERPSINFLPI---SDDKYIKLGGGLERT-QBEERKYSRDAE 122
y 126 AYRFRVAVMSERSHVMK--AFSTPTT-CSNLIGAFGLATARGNSELRSQ-----F 174
b 123 VLPAYAMLDLIG-DILRLAQTPPNLGGPLGLLRALRQGRGLAFLSRQKRDLDLF 181

```

```

Qy 175 LAPGDALLDEYDSEALKAALANFCAQSPPMSEPGTAPMVGFPAALMEVLPP----- 226
Db 182 TKSARDVLDGWFESFVKAAGF-FDAVVG-NPASPT-PGSAYVLHHTFGBVNCCKGAW 238
Qy 227 GRAVGGSGALSAAASRAMVADGATVALCGVTTSIRNSNHMT-VTTESGREVHAKVAG 285
Db 239 GHAVGGGAIQTAMAKACEAAGVEILLDAPVEAVHIDGKKAAGVQLVDGRQIMAPIVAN 298
Qy 286 CH-ILTTLLGNGGFD---RTILDHWRKIRVGGIGAV-LRLATSALPSYR---GDAT 337
Db 299 VNPALYKKLVPPSALTPTDFKAVDGYKN-----GSGTFKNVALSELPSFTCLPGKET 352
Qy 338 TRETSYSGQLIVS----DRAHLRTAAGALAGELPPRPVAVLCMGFSGIDPTIAPAGRHQV 393
Db 353 AEHQSGIVIASLDYMDAAY-RDAKGQGIS----KAPIVEMLIPLSSLDTSIAPPGQHVA 407
Qy 394 TWSQWQPYRLSGHRDWSVAEABADRIVGEMEAFAFGDTSVLDRFICTPRDISELGM 453
Db 408 SLFCQQAPELPDGRSWDDAREAADLIIDTVQWAPGFKASVGLGRMILSPDLERKFL 467
Qy 454 IGGNVMEVMSLDQMLWRPLPELSCHRVPGADGLYLTGASTHPGGVSGASGRSAARIA 513
Db 468 IGGDINHGKSLDQLWATEPLIGHASHRAPIA-GLYMOGAGTHPGGVSGNPGRNAAREI 526
Qy 514 LSD 516
Db 527 LRD 529

RESULT 10
Q98FP7 PRELIMINARY; PRT; 539 AA.
AC Q98FP7;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Phytoene dehydrogenase.
GN MUR3674.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAF303099;
RC MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shlimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
t "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003002; BAB50520.1; -.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
KW Complete proteome.
SQ SEQUENCE 539 AA; 59427 MW; 9DD78A7115A50BED CRC64;

Query Match 19.4%; Score 537.5; DB 16; Length 539;
Best Local Similarity 30.4%; Pred. No. 2.2e-25;
Matches 168; Conservative 92; Mismatches 218; Indels 75; Gaps 21;

Qy 6 DAVVCSGHNALVSAYLAREGWSVEVLEKDTVLGAVSTVERFPYKVDKGSSAHLMI 65
Db 5 DAVIIGGHNGLVCAYLAKAGLKVTVCARGVVGGAATVEFFHGFGR-NSVASTVSL 64
Qy 66 HSGIIBELGAGHGLRYIDCDPWAFAPAGTDPGPIVFHRLDTCQSIERACTKQAD 125
Db 65 PE-IMEDLELPRFLQVISYEGAVFT----RDGYLANRYRDHARREFAR-FRRDAE 118

```

```
/ 126 AYRFVAVWSRSHVMK-APSTPTGSLN-----IG-----AFGLATARGNSLSRQ 173
/ 119 AYDRYARDVTRQCRFIQPLLMRTAPDPTSFPRDITGELLYLKPKFAGLSAEEMALTRFW 178
/ 174 FLAPGDALLDEVEDSEALKAAALW---FGAQSGLPMSEPGTAPWVGFPAALMHLVLP----- 226
/ 179 TWSISD-FLDEYFEDVTKANFALSGIIGTALGP--MSPGTA-----VLLHHVGEVDGS 231
/ 227 -----GRAVGGSGALSAALASMAVDGATVAG---DGVTSIRNSNHWTVTTSREHVA 279
/ 232 VGAWYARGGCMGAVTKAALASFKASGGTIRTGAEVDHLVSRGKAK--GVVLAGEBEVYG 289
/ 280 KXVIAAGCHILTTLLDNGGDFRITLTHWR-----KIRVPGPGICAVLRATSLPS--- 331
/ 290 KLVSNADVKTFLKLV-----EKELPDIFLRVNVFKIR---GSSGKVNIALDSLPFPA 343
/ 332 -----YRGDATTRESTSGIQLVSDRAHLRTAHGAALAGELPPRAVLGMSFGSIDPT 384
/ 344 LAKDSPVYRGDMHPTDSIERMERAYDDWK-----AGWSADPFLDWVPTTLTPT 393
/ 385 LAPAGRHQVTLWSQWQVRLSGHROWASV-AEADRIVGEEMAFAPGFTSDVLDRIQT 443
/ 394 MAPGCKHMSCFVQVAPPKVG--RDWTDADRGFAESVVAQIAEYSPGFRDRIVHMEVRT 452
/ 444 PRDIESELM:GGNVHVMESLDQMLWRPLPELSGERVPGADGLYLTGASTHPGGVSG 503
/ 453 PREIEAEVLTEGNIFOGELTFDQLLENRPVPGVAYRSP-VGGLYWCSTHPGGVWG 511
/ 504 ASGRSAARIALSD 516
/ 512 APGRNAAAEILRD 524
```

RESULT 11

Q974W1

Q974W1 PRELIMINARY; PRT; 518 AA.

01-OCT-2001 (TrEMBLrel. 18, Created)

01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

Phytoene dehydrogenase related protein.

SS02422

Sulfolobus solfataricus.

Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

Sulfolobus.

NCBI\_TaxID=2287;

[1]

SEQUENCE FROM N.A.

STRAIN=ATCC 35092 / DSM 1617 / P2;

MEDLINE=21332296; PubMed=11427726;

Awatey M.J., Chan-Weiler C.C.-Y., Clausen I.G., Curtis B.A.,

De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,

Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,

Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,

Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,

Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;

"The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";

Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).

EMBL; AE006842; AAK62566.1; ..

PIR; G90413; G90413.

GO; GO:0015036; F:disulfide oxidoreductase activity; IEA.

GO; GO:0006118; P:electron transport; IEA.

InterPro; IPR000759; Adrnux\_reductase.

InterPro; IPR001327; FAD\_Pyr\_redox.

InterPro; IPR00205; NAD\_BS.

InterPro; IPR008151; Phyt dehydro.

InterPro; IPR001100; Pyr\_redox.

PRINTS; PR00419; ADXRDTASE.

PRINTS; PR00368; PADPRN.

PRINTS; PR00411; FNRDRTASII.

ProDom; PD139017; Phyt dehydro; 1.

Complete proteome.

SQ SEQUENCE 518 AA; 57944 MW; B538CF7B74ACD72B CRC64;

Query Match

Best Local Similarity 30.9%; Score 533; DB 17; Length 518;

Matches 168; Conservative 93; Mismatches 226; Indels 56; Gaps 21;

QY 5 LDVVVGGSHNALVSAALAREGMSVEVLEKDTVLGGAVSTVERFPYKYVDGRGSAHLMI 64

DB 2 IDVALIGGGHGLVTAAYLAKAGLKAVFERREIVGGASVTEELWPSIKVSTGAYVLSLL 61

QY 65 RBSGIIIEELGAGHGLRYIDCDPMAFAPAPGTPGPIVPHRDLDATQSLERACGTDA 124

DB 62 RK-ILIEELKREFGLKYLKDPGLF---LPFENGKLYIWSSELEKTKETEK-FSKNDA 116

QY 125 DAYRFFVAVW---SERSRHVMKAFSTPP---TGSNLICAFGLATARGNSLS-----RQF 174

DB 117 KNYKKWKFVDFLFAEMADFFW--LNPPPHIDEASNLINIPRG---NVNEELALSFLRTP 170

QY 175 LAPGDALLDEYFDESEALKAAAL---WFGAQSGLPMSEPGTAPWVGFPAALMHLVLP----- 226

DB 171 MODAKSLDEYFDESEALKAAALDEEDFL---RRVALKNVGVSVFXIVGYLEELPDFGNGKS 224

QY 227 ---GRAVGGSGALSAAL---ASRMADVAGATVALGDGVTISRNSNHWTVTTSREHVAH 280

DB 225 GAWGVVEGGMSEVTAQKLSAEHLGVEIYTNAEDEV--LVKNRVEGKILKNGKTINAK 282

QY 281 KVIACCHILTT-LDLLNGGDFRITLTHWRKIRVPGIGAVLRLA--TSALPSYRGDAT 337

DB 283 IIVSNADPKTTFFKLLRNALEEDFL---RRVALKNVGVSVFXIVGYLEELPDFGNGKS 338

QY 338 TRESTSGIQLVSDRAHLRTAHGAALAGELPPRAVLGMSFGSIDPTIAPAGRHQVTLWS 397

DB 339 LSPHIASELIMPENVEIKAYDDARALGYSREPLWLSINIQSSVDPTVAPPKFSFSLFG 398

QY 398 QWQPYRUSGHRDNASVAEAEADRIVGEEMAFAPGFTSDVLDRIQTPTDIESELGMIGN 457

DB 399 QYLVYD-SKRD--EMKEKTAETFEKIKEPAPNFKIKYE--VLTFLDIERRRGGINGN 453

QY 458 VMHVMSLDQMLWRPLPELSGERVPGADGLYLTGASTHPGGVSGVSGASRAIALSDS 517

DB 454 IFHLDMPDQLYFPRPLIGYSIDYTP--IRGLYLCGSHGTHPGGVGAGPNAHVKILEDL 512

QY 518 RRG 520

DB 513 RKG 515

RESULT 12

Q974W1

ID Q974W1 PRELIMINARY; PRT; 517 AA.

AC Q974W1;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein ST0549.

GN ST0549.

OS Sulfolobus tokodaii.

OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

OC Sulfolobus.

OX NCBI\_TaxID=111955;

RN [1]\_TextID=111955;

RP SEQUENCE FROM N.A.

RC STRAIN=JCM 10545 / 7;

RX MEDLINE=21456156; PubMed=11572479;

RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,

RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,

RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,

RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,

RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,

RA Oshima T., Kikuchi H.;

RT "Complete genome sequence of an aerobic thermoacidophilic

Crenarchaeon, Sulfolobus tokodaii strain7.";

RL DNA Res. 8:123-140(2001).

```
DR EMBL; AP000982; BAB5546.1; --
DR GO: 0006118; P: electron transport; IEA.
DR InterPro; IPR000759; Adnrx_reductase.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR InterPro; IPR002005; Rab_GDI_REP.
DR PRINTS; PR00419; ADXRDTASE.
DR PRINTS; PR00891; RABGDIREP.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
DR Hypothetical protein; Complete proteome.
DR SEQUENCE 517 AA; 57965 MW; E12944873A27A24 CRC64;

Query Match      19.0%; Score 525; DB 17; Length 517;
Best Local Similarity 30.1%; Pred. No. 1.3e-24;
Matches 165; Conservative 89; Mismatches 228; Indels 66; Gaps 20;

YX 1 MSAFDVAVVGSGNALVSAAYLAREGMSVEVLEKDTVLGGAVSTVERPPGYKVDGSSA 60
Yb 1 MRYIDVVIIGGNGHGVVAAYLAKGLKAVATERNVIGGATVTELPFGIKVSTASTV 60
Y 61 HLMTRHSGIIIEELGLGAGRLYIDCDPWAFAPAGTGDGFGIVFHRDLDTQCSIERACG 120
Yb 61 FSLPRKE-IIDDLRLYDFGLKVLKDPSELPVPG---QNKSIITWTSTKTKVKEIK-PS 115
Y 121 TKDADAYRRFV---AVWSERSRHVMKAFSTP---PTGSLNIGAFGLATARGNS-ELSEQ 173
Yb 116 KDRQSTKRVKILEIFSEIDFLI--LNKPINLSAEELLKLFGLNVDENVTALTIRM 173
Y 174 FLAPGDALLDEYFDSEALKAL---AMFGAQSGPMSSEPGTAPMVGFAALMHVLP-- 226
Yb 174 FFDGKSFLEDFESESSEVKALIEDSVVGTYASP--STPGTA---YVLLHNFGEVNGV 227
Y 227 ---GRAVGGSGALSAALASMAVDGATVALGCVTS--TRNSNHVTVTSGREHARK 281
Yb 228 KGAWYVEGGMGVNSALRRVABYGGAFPLNSEVDEIIVKGEAKGIKRGKIESKI 287
Y 282 VIAGCHILTT-LDLLGNGGDRITLDRHWRKIRVGPICIGAVLR--ATSALPSYRGDATT 338
Yb 288 VLSNADPKTFLKLRNAELD---EPIRKVNAKTIGVSPKINGYIEELDFGNGKSL 343
Y 339 RESTSGQLLVSDRAHURTHAGAA-----LAGELPPRAVLGMSFGIDPTIAPA 388
Yb 344 RPEHIASELIIIPSVSVYKAYLDKAKLVGSRPWLINIP-----STVDPTLAPQ 393
Y 389 GRHQVTLMSQVQYRLSGHRDWSVAEASADRIVGMEAFAPGFTDVLDRPIQTPRIE 448
Yb 394 GKPVTFIFGVYVY----SKLDKIKDKIAISIEKRVFAPNFKPYE--ILPLDIE 447
Y 449 SELGMIGNVMHVMESLDQMLRPLPELSGHRVPGADGLYLTGASTHPGGVSGASGRS 508
Yb 448 RRFQIWEGNIFHLDMTDQQLYFRPLIGYSNVRTP-IKNLYLCGSGTHPGGGVTCAPGVN 506
Y 509 AARIALSD 516
Yb 507 ASMEILKD 514

RESULT 13
8PJ01 PRELIMINARY; PRT; 539 AA.
D Q8PJ01
C Q8PJ01
T 01-OCT-2002 (TrEMBLrel. 22, Created)
T 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
T 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
E phytoene dehydrogenase.
N XAC2744.
S Xanthomonas axonopodis (pv. citri).
C Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
C Xanthomonadaceae; Xanthomonas.
X NCBI_TaxID=92829;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=306 / ATCC 13902 / XV 101;
```

```
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quesgio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.P.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Cimarotte G., Cannavan F., Cardozo J., Chambergio F., Chiapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Gregio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locall E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.P.,
RA Spinoza L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.P.,
RA Setubal J.C., Kitajima J.B.;
*Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.;
RL Nature 417:459-463(2002).
DR EMBL; AB011914; AAM37589.1; --
DR InterPro; IPR008151; Phyt_n_dehydro.
DR ProDom; PD139017; Phyt_n_dehydro; 1.
KW Complete proteome.
SQ SEQUENCE 539 AA; 57459 MW; 5E9801191CB48C3D CRC64;

Query Match      18.4%; Score 509; DB 16; Length 539;
Best Local Similarity 32.3%; Pred. No. 1.3e-23;
Matches 182; Conservative 73; Mismatches 213; Indels 96; Gaps 26;

QY 5 LDVAVGSGNALVSAAYLAREGMSVEVLEKDTVLGGAVSTVERPPGYKVDGSSAHLMI 64
Db 16 LDALIGNGHGVVAAYLAKGRVLVLEARVVGGAATVEEFHGFGR-NSVAATVSL 74
QY 65 RHSGIIEELGLGAGRLYIDCDPWAFAP-----PAGTGD---QGFIVFHRDLDTAQ 113
Db 75 LQPRVIALDLALERHGLRV-----PRINNPLPLPNGDYLLAGAG-----RTAQ 118
QY 114 SIERACTKDADAYRRFVAVWSERSRHVMKAFSTPP-----TGSNLIGATGGLATA----- 164
Db 119 DVAR-PSARDAALPALYEA-RLEQLADVLRLALALQPPDVTGGWLQALPOLWRACKLGL 176
QY 165 --RG-----NSLSRQFLAPGDALDEYFDSEALKALAW---FGAQSGPMSSEPGTAPM 214
Db 177 QLQGLPLTLRQELLDLTISAAEYLDWTFESAIPKALFGFDGIVGNVYASP--HTPGTA-- 232
QY 215 VGFPAALMHVLP-----GRAVGGSGALSAALASMAVDGATVALGCVTSIRNSNH 266
Db 233 --YVLLHHVFGQSGNGYKAGWGAIGMGAITQMAASAEYGAQLTGCASERVIVEQGH 290
QY 267 WT-VTTESGREVHARKVIAGCH-ILTTDLGNGGDRITLDRHWRKI---RVGPGIGAV 321
Db 291 AVGVVITAGETVBARAVANVNPVKLYQLMRPDDVPATT-----REIAHYRCGSGTFR- 345
QY 322 LRLATGALPSY-----RGDAITRESTSGQLLV-----DRAHLRTAHGAALAGELPPRPA 372
Db 346 MNVALSELDPFAALPGND-----HLTAGIILAPSLDYMDRAW-----HDAFAFG-WSRBEV 396
QY 373 VLGMSFGSIDPTIAPAGRHQVTLMSQVQYRLSGHRDWSVAEASADRIVGMEAFAPGFG 432
Db 397 VEILLPSTLDDTLAPGQHVASFQCHVAPQLPDGRRDDHREVDLMIATVRYAPGF 456
QY 433 TDSVLDRIQTTPRIESELGMIGNVMHVMESLDQMLRPLPELSGHR--VPGADGLYL 490
Db 457 AASVLGRQVLSPLDLERMFGLVGGDIFHGALSINQLFSGARPLLGQGVYRGAVP---GLYL 513
QY 491 TGASTHPGGVSGASGRSAARIAL 514
Db 514 CGSGTHPGGGVTCAPGHNAARVIL 537

RESULT 14
Q89FW3
ID Q89FW3 PRELIMINARY; PRT; 545 AA.
```



Q39FW3;  
01-JUN-2003 (TReMBLrel. 24, Created)  
01-JUN-2003 (TReMBLrel. 24, Last sequence update)  
01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
BLR6586 protein.  
Bradyrhizobium japonicum.  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Bradyrhizobiaceae; Bradyrhizobium.  
NCBI\_TaxID=375;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=USDA 110;  
MEDLINE=2248498; PubMed=12597275;  
Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,  
Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,  
Kohara M., Matsumoto M., Shimo S., Tsurucka H., Wada T., Yamada M.,  
Tabata S.;  
"Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110.";  
DNA Res. 9:189-197(2002).  
EMBL; AF005959; BAC51851.1; -.  
GO: 0016491; F:oxidoreductase activity; IEA.  
GO: 0006118; P:electron transport; IEA.  
InterPro: IPR000205; NAD BS.  
PRINTS: PR000103; Pyridine\_redox\_2.  
Complete proteome. PNDRTASEII.  
SEQUENCE 545 AA; 59480 MW; 0B2DDC6F39C0B783 CRC64;  
Query Match 18.0%; Score 497; DB 16; Length 545;  
Best Local Similarity 30.5%; Pred. No. 7.4e-23;  
Matches 174; Conservative 77; Mismatches 217; Indels 102; Gaps 25;  
5 DAVVVGSGNALVSAYLAREGNSVEVLEKDTVLGGAVSTVERFPGYKVGSGSAHMR 65  
14 DVVVIAGGNGTCAAYLAMAGLRVVRVERKVGGAATFEFHPGFR-NSVAYTVSL 72  
66 HSGIIEELGLAGHLYIDCPWAFAPAPGT---DGPGLVFHRDLATCQSIERAGTK 122  
73 NPQVIRDLGLAQGLRIVERAQNPLPADGSYLLTGE---RTKASVAR-LSAH 123  
123 DADAYRRFVAVWSESRH-----VMKAP--STPTGSLNIGAF----- 159  
124 DADALDGF-----SRELEDIADVLROFVLRAFP---NLIDGFGTNAIREAVNALQSAN 173  
160 ---GLATARGNSLSRQFLAPGDALDEYDSEALKAAALWFGAQSGPPMS--EPGTAPM 214  
174 ILRGITLQSRSLD-LFTRSGEMLDERFEDLVK-ALPGFDAIVGNYPYAGSA-- 229  
215 VGFALMHVLP-----GRAVGGSGALSAAALASRMAYDVGATVALGDGVT-IRNSN 265  
230 ---YVMLHAFGEVANGKGVWGHGALGEMGATQAWARAAQGRGVAIDTDAGREVIVEROR 287  
266 HWTVTTEGSRVHARKVIAGCH-----ILTDLLNGGDFDRTLDHWRKTRVPGIG 319  
288 AVGVLENGATIRAKYVAANVPKLLVTRISAADALPKDFLAR--IHHWKN-----GSG 339  
320 AV-LRLATSLPSYR-----GDATTRESTSLQLLV-----DRAHL-RTAHGAALAGELP 368  
340 TFRMVALDLRPSFTALPDGCD---HLTSGIILAPSLGYMDRAFLDARAGHS----- 389  
369 PRPAVLGMSFGIDPTAPAGRHQVTLWSQVRLSGHRDNASVAEADRIVGMBAF 428  
390 REPVEVLEIPSTLDDTLAPGKRVASLFCQHVAPFLPDSRSNDHDEVDLMIATVDSY 449  
429 APGETSDVLRPIOTPRDISELGICGNVHVMESLDQMLWRPLPELSGHRVPGADGL 488  
450 APGASSVLGEIILSPDLREQFLLGSDIFHGALTLNQLFSARPMLGHADYRGP-LKGL 508  
489 YLTGASTHPGGVSGASGRSAARIALSDSR 518  
509 YHCGSGAHPGGVGTGAPGHNAAQAAILRDHR 538

## RESULT 15

Q98BS8 PRELIMINARY; PRT; 524 AA.  
ID Q98BS8  
AC Q98BS8;  
DT 01-OCT-2001 (TReMBLrel. 18, Created)  
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)  
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)  
DE Probable dehydrogenase.  
GN ML15443.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
DR EMBL; AF003006; BABS1894.1; -.  
DR InterPro: IPR008151; Phytin dehydro.  
DR ProDom: PD319017; Phytin dehydro; 1.  
KW Complete proteome.  
SQ SEQUENCE 524 AA; 55519 MW; 65B9A7400C90B282 CRC64;  
Query Match 16.9%; Score 469; DB 16; Length 524;  
Best Local Similarity 29.8%; Pred. No. 3.9e-21;  
Matches 161; Conservative 81; Mismatches 238; Indels 60; Gaps 22;  
QY 1 MSAFDVAVVGSGNALVSAYLAREGNSVEVLEKDTVLGGAVSTVE-RFGYKVDGSS 59  
Db 1 MSERF-DALFVGAGHNSLCAALALKGWKTGIFERSSTIGGAVQTREFTLPGFHRDFGM 59  
QY 60 AHLMTRHSGIIEELG--LGAHGLRY---IDCDPWAFAPAPGTDPGIVFHRDLATCQS 114  
Db 60 NLSLFAGSAFHRKYANELKTQGLEFAPVADC---FASAPP--DGRWFGVNDLEKTASR 113  
QY 115 TERAGTGDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGGLATARGNS---ELS 171  
Db 114 M-AAFSAADAATWRKLVAAFPGEAEHLFRLIGSPMSARALAGTANLWRKKGAVAGALDTG 172  
QY 172 PQFLAPGDALLDEYDSEALKAAALWFGAQSGPPMSBEGTAPMVGFALMHVLP----- 226  
Db 173 RLLSSPRTWLEETFESEPPYRATLATWGMH-----DFAPDIAGGAVFPYLESMANQS 225  
QY 227 -GRAVGGSGA--LSAALASRMAYDVGATVALGDGVTISIRNSNHWVTTSBGRVH--ARK 281  
Db 226 FQWVLGKGADTIIRALAGMVTSGAGKIVTCAEYSEITVSNKATGVRLTSGETHATKA 285  
QY 282 VIAGCHILTTLDL-----GNNGDFDRTLDHWRKIRVGPICIGAVLRLATSLPSVGDGA 336  
Db 286 VIAGVAPKALTGTLPGSGNGAGFDTA-----MKFYAFGT-WMHLALDELPGWRAGS 339  
QY 337 TTRSTSGQLLVSDRAHLRTAHGAALAGELPPRPAYLVGMSFGIDPTIAPAGRHQVTLW 396  
Db 340 ELRQ-FAYVTLSPSLDAMSR-T-YQAMAGMLPQEPVLVVGQPTAIDPSRAPQGH--VLW 395  
QY 397 SQWQ--PVRLSGH-----RDWASVAEAEADRIVGMBAFAPGFTDVLDRITQPRDI 447  
Db 396 VQWRMLPAETIGDAGGKIAPAHMDQVADVAERVLDTIETIYAPGLRSKILGRSVFSDIL 455  
QY 448 ESE-LGMIGGNVHVMESLDQMLWRPLPELSGHRVPGADGLYLTGASTHPGGVSGASG 506  
Db 456 ERENPNLVGGDQVCGSHLQNFILFRPARGVACWNTP-VGNLHLTGAATWPGAGTGAAG 514

Search completed: February 29, 2004, 14:51:13  
Job time : 48.746 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

in on: February 29, 2004, 14:27:18 ; Search time 8.96629 Seconds  
(without alignments)  
3089.496 Million cell updates/sec

file: US-09-941-947A-38

Effect score: 2768  
Sequence: 1 MSALFDDVVGSGHNLVSA.....ALSDSRGKASQWRRSSRS 532

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Aligned: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	327	11.8	469	1 P49_STRLI	P06108 streptomyc
2	303	10.9	535	1 Y997_MYCTU	Q10555 mycobacteri
3	285	10.3	508	1 CRTI_STRSE	P54971 streptomyc
4	268	9.7	507	1 CRTI_STRGR	P54981 streptomyc
5	226	8.2	495	1 CRTD_RHOSH	Q01671 rhodobacter
6	224	8.1	621	1 CRTI_CERNC	P48537 cercospora
7	203.5	7.5	595	1 CRTI_NEUCR	P21334 neurospora
8	203.5	7.4	518	1 CRTI_RHOSH	P54980 rhodobacter
9	191.5	6.9	492	1 CRTI_PANAN	P21685 pantoea ana
10	189.5	6.8	529	1 CRTI_MYXXA	Q02861 myxococcus
11	179.5	6.5	492	1 CRTI_ERWHE	P22871 erwinia her
12	176	6.4	494	1 CRTD_RHOCA	P17059 rhodobacter
13	176	6.4	583	1 CRTI_PHYEL	P54982 phycomyces
14	174.5	6.3	501	1 CRTI_AGRAU	P54978 agrobacteri
15	174	6.3	524	1 CRTI_RHOCA	P17054 rhodobacter
16	158.5	5.7	517	1 CRTJ_MYXXA	P54979 myxococcus
17	139.5	5.0	471	1 PPOX_MYXXA	P56601 myxococcus
18	134	4.8	507	1 MQO2_PSEAE	Q9hvf1 pseudomonas
19	131.5	4.8	548	1 PPOC_TORAC	Q24163 nicotiana t
20	130	4.7	471	1 Y782_SVNY3	Q55629 synchocyst
21	129	4.7	452	1 PPOX_MYCTU	Q53303 mycobacteri
22	129	4.7	474	1 DLDH_HALVO	Q04829 halobacteri
23	128.5	4.6	555	1 DLHD_STRHA	Q05355 streptomyc
24	127.5	4.6	469	1 DLHD_CHLVI	O50311 chlorobium
25	127	4.6	3590	1 PHAB_BORPE	P12255 bordetella
26	125	4.5	1001	1 ORKI_DROME	Q94526 drosophila
27	124	4.5	527	1 PPOX_PROFR	Q32434 propionibac
28	121	4.4	417	1 Y4AB_RHLSN	P55349 rhizobium s
29	119.5	4.3	529	1 SP15_STRGR	P19471 streptomyc
30	119	4.3	395	1 KIME_MOUSE	Q9r008 mus musculu
31	119	4.3	567	1 FIG1_HUMAN	Q967q9 homo sapien
32	118.5	4.3	4303	1 PKD1_HUMAN	P98161 homo sapien
33	118.5	4.3	5262	1 MLL2_HUMAN	O14686 homo sapien

## ALIGNMENTS

### RESULT 1

ID	P49_STRLI	STANDARD;	PRT;	469 AA.
AC	P06108;			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JAN-1988 (Rel. 06, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	P49 protein.			
GN	P49.			
OS	Streptomycetes lividans.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycineae; Streptomycetaceae; Streptomycetes.			
OX	NCBI_TaxID=1916;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=66 / 1326;			
RX	MEDLINE=87231086; PubMed=3453116;			
RA	Burnett W.V., Henner J., Eckhardt T.;			
RT	"The nucleotide sequence of the gene coding for XP55, a major secreted protein from Streptomycetes lividans.";			
RL	Nucleic Acids Res. 15:3926-3926(1987).			

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; Y00142; CAA68336.1; -;  
CC PIR; S09189; S09189.  
CC InterPro; IPR000759; Adrnkx reductase.  
CC InterPro; IPR000205; NAD BS-  
CC InterPro; IPR008151; Phytin dehydro.  
CC PRINTS; PR00419; ADXRDASE.  
CC Prodom; PDI39017; Phytin dehydro; 1.  
CC SEQUENCE 469 AA; 49740 MW; C2AE7533C7C701CB CRC64;

Query Match 11.8%; Score 327; DB 1; Length 469;  
Best Local Similarity 27.6%; Pred. No. 8.6e-15;  
Matches 147; Conservative 64; Mismatches 223; Indels 98; Gaps 25;

QY	5	LDVVVSGHNLVSAAYLARSGNSVELEKOTVGGAVSTVE-RFPGYKVDGSSAHLM 63
DB	2	LDVVVVGAGPGLTAAYVELARRGFPVAVFEAQGTGGGARTTELTLPGRPHDPCSAHPL 61
QY	64	IRHSGIIEELGLGAHGLRYICDWPAPAPGTDGPGIVFHRDLDTATCOSIERRCTKD 123
DB	62	GINPAPRGPLERYGLENLH-PGLPMHFFP--DSAAVLSRVGETAASF-----GARD 114
QY	124	ADAYR-----RFVAVWSRSRHVAFSTPTPTGNLIGAFGLATARGNSLSQFLA--P 177
DB	115	AGPYRRLIERFLPRWDTLARDFM---SLPLT-----ALPRDPVTLARFGLVGLP 160
QY	178	GDALLDFYDFSEALKKALAWFGAQSGPPSEPCPTAPMVGFALMHVLPFGRAVGGSGALS 237

```

161 PSTWLMERFDEKAKTLEAGLVAVH-----MAPLGGFAT-----GAILGVFALA 204
238 AALASMAVVGATVALGDDGVTSTRNSNHWVTYTESGRVHARKVI-----AGCHIL 289
205 AHARGWVARGGSSQSDAUYLK-----DUGGAVHTDYEVRKLDLPPARAYVL 255
290 TT-----LELLGNGGDFRTTLDHWRKIRVGPQIGAVLRATLSALPSYRGDATTRESTS- 343
256 DTSPTALARIAGLG-----SHY-ANYRVPSPFKIDVALDGPVP-----WTAEPERSA 302
344 GLQLLYSDRAHLATA-HGALAGELPPRAVLGMSPGIDPTIAPAGRHQVTLNSQWQPY 402
303 GTVOIGADSTEIGAAHLHAPSGLDRAPERFLITVQPSVADPTAPAGRH--VFWAY----- 356
403 RLSCH--RDWASVAEAEADRIVGEMAEAFAGFETSDVLDRIQTPRODIES-ELGMITGNNVM 459
357 ---GHVPNGWNGDL-----TDAMEQLERFAPAGFRDRVLARATAGPAELARNAVYGGDIS 410
460 HVMSLSDOMLWRPLBELSGHRVPGADGLYLGTASTHPGGGVSAGSRSAR 511
411 SGAVSGLQLLL-RPKISLPFYSTP-RPAVFICSSATPPGPGVHGMSGHNAK 460

```

RESULT 2

897 MYCTU STANDARD; PRT; 535 AA.

Q10555;  
01-OCT-1996 (Rel. 34, Created)  
01-OCT-1996 (Rel. 34, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
E Hypothetical protein RV0897C/MT0921/MB0921C.  
RV0897C OR MT0921 OR MTCY31.25C OR MB0921C.  
Mycobacterium tuberculosis, and  
Mycobacterium bovis.  
S Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
C Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
X NCBI\_TaxID=1773, 1765;  
N [1]  
P SEQUENCE FROM N.A.  
C SPECIES=M.tuberculosis; STRAIN=H37RV;  
X MEDLINE=98295987; PubMed=9634230;  
A Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
A Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,  
A Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
A Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
A Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,  
A Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
A Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
A Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
T "Deciphering the biology of Mycobacterium tuberculosis from the  
complete genome sequence.";  
L Nature 393:537-544(1998).  
N [2]  
P SEQUENCE FROM N.A.  
C SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;  
X MEDLINE=22206494; PubMed=12218036;  
A Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
A Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
A Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,  
A Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
A Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
T "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains.";  
L J. Bacteriol. 184:5479-5490(2002).  
N [3]  
P SEQUENCE FROM N.A.  
C SPECIES=M.bovis; STRAIN=AF2122/97;  
X MEDLINE=22709107; PubMed=12788972;  
A Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,  
A Pryor M., Duthoy S., Grondin S., Lacroix C., Morsempe C., Simon S.,  
A Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
A Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;

\*The complete genome sequence of Mycobacterium bovis.\*;  
Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; Z73101; CAA97372.1; -;  
DR EXBL; AE006979; AAK45167.1; -;  
DR EMBL; BX248337; CAD93782.1; -;  
DR PIR; F70782; F70782.  
DR TIGR; MT0921; -;  
DR Tuberculist; RV0897C; -;  
DR InterPro; IPR001613; Amineoxid\_fl.  
DR InterPro; IPR00205; NAD\_BS.  
DR InterPro; IPR008151; Phyt\_n\_dehydro.  
DR PRINTS; PR00757; AMINEOXDASEF.  
DR ProDom; PD139017; Phyt\_n\_dehydro; 1.  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSMEM 7 27 POTENTIAL.  
FT TRANSMEM 509 529 POTENTIAL.  
SQ SEQUENCE 535 AA; 56233 MW; CDAAA92094DCB7D6 CRC64;  
Query Match 10.9%; Score 303; DB 1; Length 535;  
Best Local Similarity 24.4%; Pred. No. 4.1e-13;  
Matches 147; Conservative 74; Mismatches 214; Indels 168; Gaps 25;  
QY 6 DAVVCSGNAVSAYLAREGMSVELEKDTVLGAVSTVERFPGYKVDROSSAELM-I 64  
DB 9 DVVVGGHNGLLVAAAYLARAGLRVLLERLAQTGGAASVIOAFDGEVVALSRYSL 68  
QY 65 RHSGIIIEELGAHGLRYDCDFWAFAPAPGTG-PGIVFHRDLDTACQSIERACGKTD 123  
DB 69 LPSRIVADLGAPVRLAR---RPFSSYTPAPATAGRSGLLIGTGPRAAHLAAGAAPD 124  
QY 124 ADAYRFP-----VAVW-----SERSH-----VMAKFSPTPTGSNL 154  
DB 125 AGGFAAFYRCRLVTLARLWFTLIEPLRTREQARDIVYEGHEAAAAAQAQVDEPIGHAI 184  
QY 155 IGAFGGLATARGNSLSROFLAPGDALLDEVDSALKAAALWFAQSGPPMSEBGTAPM 214  
DB 185 AGAVA-----NDLLRGVIAT-DALIGTP-----ARMHEBPLMQN 217  
QY 215 VGFPAALMEVLPGRVAV-----CGSGALSAALASRVAVDQATVALG-----DGVTSIR 261  
DB 218 ICF-LYHLVGGGTGVNHPVIGMGSVTSALATAAAREGAEIVTGADYFALDPOG--TVR 273  
QY 262 RNSNHWVTVTESGRE--VHARKVIAGCHILATLIDLNGSGDFDRTTLDHWRKIRVGPQIG 319  
DB 274 YESD-----GSDGAELVNRGFRVTVGTPAVLASLLG-----EPVVALAPGAQ 316  
QY 320 AVLRATLSALPSYRGDATT-RESTSGLQLLYSDRAHLRTAHGAALAGELPPRAVLGNSF 378  
DB 317 VKYNMVMVRELRLRDRDSVTPOQAFAGTFHVNWTSQDLDAAYSQAASGLRDPDLPCEAYCH 376  
QY 379 SGIDPTI-----APAGRHQVTLWSQVPRLSGHRDVAEAEADRIVGEMAEAPAGPT 433  
DB 377 SLTDSILSARLDAGAQTLLTVFGLHTPSHVPD-----TEGLAERTAAVLA---SL 426  
QY 434 DSVLDRFIQ-----TPRDISELGMIGCNVNVHMSLDQMMLRPLPEL 477  
DB 427 NSVLAEPIDVLTWDAQSKPCIETTTLDLQRTLQRTGCTGNI FHGALS-----W----- 474  
QY 478 SGHRVPGADG-----LYLTGASTHPGGGVSAGSRSARIALS--D 516  
DB 475 -----PFADNDPDLTPARQWGVATDHERIMLCUGSARRRGVAGSICGHNAAVLA 529  
QY 517 SRR 519  
|||

b 530 SRR 532

```
RESULT 3
R1_STRS
C RTI_STRS STANDARD; PRT; 508 AA.
1 01-OCT-1996 (Rel. 34, Created)
2 01-OCT-1996 (Rel. 34, Last sequence update)
3 28-FEB-2003 (Rel. 41, Last annotation update)
4 Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
5 CRTI.
6 Streptomyces setonii.
7 Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;
8 Streptomycineae; Streptomycetaceae; Streptomyces.
9 NCBI_TaxID=38315;
10 [1]_SEQUENCE FROM N.A.
11 STRAIN=ISP 5395;
12 Hoshi K.;
13 Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
14 -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
15 intermediary of phytofluene by the symmetrical introduction of two
16 double bonds at the C-11 and C-11' positions of phytoene.
17 -!- COFACTOR: FAD (Probable).
18 -!- PATHWAY: Carotenoid biosynthesis.
19 -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
20
21 This SWISS-PROT entry is copyright. It is produced through a collaboration
22 between the Swiss Institute of Bioinformatics and the EMBL outstation
23 the European Bioinformatics Institute. There are no restrictions on its
24 use by non-profit institutions as long as its content is in no way
25 modified and this statement is not removed. Usage by and for commercial
26 entities requires a license agreement (See http://www.isb-sib.ch/announce/
27 or send an email to license@isb-sib.ch).
28
29 EMBL; D55723; BAA09537.1;
30 InterPro: IPR000759; Adrnx reductase.
31 InterPro: IPR002937; Amino oxidase.
32 InterPro: IPR008150; Bac phytoene_dh.
33 InterPro: IPR000205; NAD_BS.
34 InterPro: IPR008151; Phytin dehydro.
35 Pfam: PF01593; Amino oxidase; 1.
36 PRINTS; PR00419; ADXRD7ASE.
37 ProDom; PD133017; Phytin dehydro; 1.
38 PROSITE; PS00982; PHYTOENE_DH; 1.
39 Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
40 NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
41 SEQUENCE 508 AA; 54610 MW; 4B6DEF076D51CB5 CRC64;
42
43 Query Match 10.3%; Score 285; DB 1; Length 508;
44 Best Local Similarity 25.3%; Pred. No. 6.3e-12;
45 Matches 143; Conservative 64; Mismatches 208; Indels 150; Gaps 21;
46
47 8 VVGGSHNALVSAAYLAREGMSVEVLEKDTVLGAVSTVERFPGYKVDGSSAHLIRHS 67
48 12 VVVGAGLAGLAALHLGAGRSVTVVSEQVPGGRAGLLET-DGFRVDTGPTVLTN--P 67
49
50 68 GLIEE--LGLG--AGGLRVIDCPWAFAPAGTDPGPGVIFHRDLATCQSIERACGTK 122
51 68 DLVEEAFAAAGEPMWADLEILRLDP---AYRAFADGSLQDVTGDAAMEAAVEQFAGAR 124
52
53 123 DADAYRRFVAWVSRSRHHVMAKFTPTPTGSLNLCAGFGLATARGNSLSQFPLAPGDALL 182
54 125 QAVGYR-LRIWLELRVQ-----NRRF-----I 148
55
56 183 DEYFDS-----ALKALAWFG---AOSGPPMSEPGTAPMWGFAALMHVLPGRGV- 230
57 149 DANFSDSPQLVHPDLARLAALGGFCRLDARIGHFVSDERLRRVFSFQALYAGVPPARALA 208
58
59 231 -----GGSGALSALASRMAVDGATVALGDGVTISIRNSNHWHTVITE 272
60
61 209 AYAVIAYMDTVAGVYFFRGGGHHALPRWADAADADAGASFYQSVTRLSRSDRVTAUVT 268
```

## RESULT 4

```
CRTI_STRS
ID CRTI_STRS STANDARD; PRT; 507 AA.
AC P54981; P7247;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
GN CRTI OR CRPE.
OS Streptomyces griseus.
OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
XN [1]_SEQUENCE FROM N.A.
RP STRAIN=JA933;
RC MEDLINE=97074881; PubMed=9917308;
RA Schumann G., Nurnberger H., Sandmann G., Kruegel H.J.;
RT "Activation and analysis of cryptic crt genes for carotenoid
RT biosynthesis from Streptomyces griseus."
RL Mol. Gen. Genet. 252:658-666(1996).
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
CC intermediary of phytofluene by the symmetrical introduction of two
CC double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COFACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L37405; AAA91950.1;
CC EMBL; X95596; CAA64850.1;
CC InterPro: IPR000759; Adrnx reductase.
CC InterPro: IPR002937; Amino oxidase.
CC InterPro: IPR008150; Bac phytoene_dh.
CC InterPro: IPR000205; NAD_BS.
CC InterPro: IPR008151; Phytin dehydro.
CC Pfam; PF01593; Amino oxidase; 1.
CC PRINTS; PR00419; ADXRD7ASE.
CC ProDom; PD133017; Phytin dehydro; 1.
CC PROSITE; PS00982; PHYTOENE_DH; 1.
CC Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
CC NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).
CC SEQUENCE 507 AA; 54509 MW; FBB97F7FE696B2AC CRC64;
```

```

Query March          9.7%; Score 268; DB 1; Length 507;
Besc Local Similarity 25.0%; Pred. No. 8.6e-11;
Matches 139; Conservative 65; Mismatches 217; Indels 136; Gaps 21;

      8 VVSGHNALVSAAYLAREGWSVEVLEKDFVLGAGVSTVERPPGYKVDGRGSAHLIRHS 67
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
      12 VVVGAGLAGLAAGLHLGAGRRVTVVEREDVPGRAGLLES--GGRIFTGPTVLTM--p 67
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      68 GILBELGLGAGHLYID-CDPMAFAPP--APCTDGPGVTFHEDLDATCQSTERACGTKA 124
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      68 DLVED-AFAAVGERMADRLELILAPAYRRPADSGQLDVHTDGLAMGAAYEEFAGARQA 126
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      125 DAYRFVAVMWSERSRHVMKAFSTPTPGSNLIGAFGLGATARGNSELSQFLAPGDALLDE 184
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      127 VGYAR-LRIWLRLRYVO-----MRRF-----IDT 150
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      185 YFDSE-----ALKAALAWFG--AQSGPPMWSBPGTAPVGVFAALMHVLPGRGV-- 230
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      151 NFDSPLOLAHPDLARLAALGGFGLDARIGHFVSDERLRRVFSFOALYAGVPPARALAAY 210
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      231 -----GGSGALSAAALASRAVGDATVAGLDGVTSIRNSNHWTVTTTSG 274
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      211 AVIAYMTVAGVFPFGMGHAPRAWDAASDAGAVLRYGQVTRLSRSGDRVAVVTDQ 270
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      275 REVEARKVIAGCHILATLIDLNGGQFDRITLDHWRRKIRVGPICAVLRATLSALPSVRG 334
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      271 EHIPCDAVLTDPDLVPSYRLIG-----RT-----PERELPLRHSPSAVILHTG 313
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      335 DATTRSTSGLLQVSDRAHLRTAHGAA-----LAGELPPPAVLGMSFGIDPTI 385
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      314 TDRP-----WPDLAHTTISFGAAWKNTHLTRGLRMSDPSELLITRPTATDPSL 363
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      386 APAGRHOVTLWSQWPYRLS-----GHRDMSVAEAEADRVIGEME-AFAPGPTDS 435
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      364 APPGKH-----LHYVLAPCNTVEVPGVREWEELGPRYDELLABELERREMPGLGAA 415
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      436 VLDFEPIOTPROISEELGKMGVNMVHENSILDQMLWRPLPELSGHRVPGADGLYTGAST 495
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      416 IESEGLTVPDNTAQ--GHAAGTFFSVAHTFPQTGGPFR-----RMLVRGTVNAVLAGCGT 469
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      496 HPGGVGSGA--SGRSAA 510
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      470 TPGVGVPVTLISGKLA 486
      1b      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 5

RESULT 5  
CETD RHOSH

CRTD\_RHOSH  
ID\_CPTD B

ID	CRTD_R
1	001571

AC Q01671

DT 01-JUL

16-OCT 1951

28-FEB-82

21 28-FEB  
2E Methox

EN  
METHUEN  
CRTD.

EN CRTD.  
SC Rhodob

Rhodospirillum rubrum  
Rhodospirillum rubrum

Bacter  
Phage

Rhodolite

NCBI\_T

NR [1] .

SEQUENCE

2C STRAIN

SC  
SX

Garj E

3A	Garl E
3T	"Nucle

<sup>3</sup>T "Nucle  
<sup>3</sup>T From B

from R

carotenoids

FEMS M

[2]

dr  
SEQUENCES

SEQUENCE  
STRAINING

IL  
SIRAIN  
MEDLIN

XX MEDLIN  
LA Lapa H

Lang H  
"Compl

T "Comp1

```

assignment of the carotenoid biosynthesis pathway of Rhodobacter
sphaeroides";
J. Bacteriol. 177:2064-2073(1995).
-!- FUNCTION: CONVERTS HYDROXYNEUROSPORENE TO DEMETHYLSPHEROIDEONE OR
METHOXYNEUROSPORENE TO SPHEROIDEONE.
-!- COFACTOR: FAD (Probable).
-!- PATHWAY: Carotenoid and chlorophyll biosynthesis.
-!- SIMILARITY: Belongs to the phytylene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X63204; CAA44886.1; -.
EMBL; AJ010302; CAB38743.1; ALT_SEQ.
PIR; S23633; S23633.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; Bac.phytoene_dh.
InterPro; IPR000205; NAD_B5.
Pfam; PF01593; Amino oxidase; 1.
PROSITE; PS00982; PHYTOENE_DH; 1.
Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
Oxidoreductase; FAD; Flavoprotein; NAD.
NP_BIND 9 42 FAD (ADP PART) (POTENTIAL).
CONFLICT 115 115 A -> G (IN REF. 2).
CONFLICT 157 157 T -> P (IN REF. 2).
CONFLICT 170 170 L -> M (IN REF. 2).
CONFLICT 273 273 L -> I (IN REF. 2).
CONFLICT 440 443 PHGA -> ATGP (IN REF. 1).
SEQUENCE 495 AA; 52900 MW; 62227931415B253E CRC64;

Query Match 8.2%; Score 226; DB 1; Length 495;
Best Local Similarity 26.2%; Pred.No.5.e=08;
Matches 146; Conservative 54; Mismatches 233; Indels 124; Gaps 29;

QY 8 VVVGSGHNALVAAYLAREGWSVEVLKDTVLGGAVSTVERFPCKYKVDRGSSAHLIRH- 66
Db 9 VVVGNGGLASRLARRAGCEVTLLERAPGGRMETLPSVAG-PVDAQPTV-LI.REV 66
QY 67 -SGIIIEELGLGA-HGLRYIDCDWFAPAPP-----GTGPGIVFHRRDDATCQSIRA 118
Db 67 FDIIFVCQGKLDEHLLTL-----QPOLLARHWLDGSTLDLTLEANEVVEAAVF 117
QY 119 CGTKDADAYERFVAVWSERSRHVMKAFSTPTGSNLICA---PGGLATARGNSBELSRQF 174
Db 118 AGAREAPAFRRP-----HDLSRLYDAFORP-----MKRAAPDLRALATGALKAPRTWPA 168
QY 175 LAEPG----DALLDEYFDSEAKLAALNFWAQ-SGPFMSPEPTAPMGVPAALMHVLPFGRAV 230
Db 169 LLPQMTLDRLRLFPRDRRLRLQFCRYATVYGTFYGAAGVLALIMAA-----EARGV 221
QY 231 -----GSGGALSALAARMV-D-GATVALGDGVTSIRNSNHEWT-VTTESGREVHARKVI- 283
Db 222 WAIEGGMRHALALA-RLADDQGVRLRYGATPVAGILRQGGREPTGVQADGRTLPAQHIVF 280
QY 284 -----ACCHITLLDLLGNQGFDRITLLDHWRKRVGPGIGAVLRILATSALPVSRGD 335
Db 281 NGDPALLACC-----LGDPQDAPVED-----RIHP-----RSLSANWYSYAR 320
QY 336 ATTRESTGLQLL----VSDRAHLRTAHGAALAGELPPRAPVLGMFSFGIDPTIAPAGRHQ 392
Db 321 A-----SGPPLVHHNVFFADDFREFGFIAGQM-PEDATLYICASDRGGGLPDGPER 373
QY 393 VTLSQWQVRLSGHRDNASVAEABD-----IVGEMEAFAP-GFTDSVLDRFI 441
Db 374 FEIMNGPPORPAKEDFPAQCBSRTFDRLRQFGITFDVPGETSLTPAGSFASL----- 427
QY 442 QTPRDIESIELMIGNVHVMEVMSLDOMMLWRPLPELSGHRVPGADGLYLTCGATHPGGV 501

```

```

428 -----FPASQSIYGLSPHGALA-----SLKRLPARTA-----LPGLWLAGGGAHPGAGV 472
502 --SGASGRSAARIALSD 516
473 FPAALSGRHAAREAILAD 489

RESULT 6
TI_CERNIC
P48537;
01-FEB-1996 (Rel. 33, Created)
01-FEB-1996 (Rel. 33, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
PDB1.
Cercospora nicotianae.
Eukaryota; Fungi; Ascomycota; Pezizomycotina;
Dothideomycetes et Chaetothyriomycetes incertae sedis;
Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cercospora.
NCBI_TaxID=29003;
[1]
SEQUENCE FROM N.A.
STRAIN=ATCC 18366;
MEDLINE=94368091; PubMed=8085820;
Brenshafft M., Daub M.E.;
"Isolation, sequence, and characterization of the Cercospora
nicotianae phytoene dehydrogenase gene.";
Appl. Environ. Microbiol. 60:2766-2771(1994).
-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
-!- COPACTOR: FAD (Probable).
-!- PATHWAY: Carotenoid biosynthesis.
-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U03903; AAB86988.1; -.
PIR; T48646; T48646.
InterPro; IPR008150; Bac_phytoene_dh.
InterPro; IPR008151; Phyt_n dehydro.
PRODOM; PD139017; Phyt_n dehydro; 2.
PROSITE; PS00982; PHYTOENE_DH; 1.
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;
Transmembrane.
NP BIND 11 44 FAD (ADP PART) (POTENTIAL).
TRANSMEM 536 556 POTENTIAL.
SEQUENCE 621 AA; 69529 MW; 61B9EA7784963CA8 CRC64;

Query Match
Best Local Similarity 22.1%; Pred. No. 9.9e-08;
Matches 129; Conservative 73; Mismatches 245; Indels 136; Gaps 22;

7 AVVVGSHALVGAAYLARGSGVSEVLEKDTVLGAVSTVERPPGVKVDKRSASHLMITH 66
10 AIVLGSGVGVSTAAALARGHFWTVLERNNFTGRCSLIHH-EGYRPPQGSLLLL--- 65
67 SGHIE-----ELG--LGAHSLRYIDCDPMFAPAPPAGTGDGPGIVFHRDLDTACQSIERACG 120
66 PGLFHRFTFAELGTSLEQGVKLLKCPNYM---IHFDGKEKFTLSDDLVMKTEVEKWEK 122
121 TKDADATRRVAVNSRSRHVKAFSTPPFGSNLIGAFGLATARGNSLSQFLA--PG 178
123 ---KEGTRYLEFLKSHGHYELS-----VREYLLRNFEGL-TAMLRPEFLRHLLQLHPF 173
179 DAL---LDEYFDSEALKAALAWGAQSG-PPMSEPGTAPMWGFALMHVLPGRVAVGGSG 234

```

```

Db 174 ESIWTRACKYFWTERLRVRVFTFGSMYMGMSFPDAPGTVSLQLQYTELAEGL--WYPVGGFH 231
QY 235 ALSALASMAVDGATVALGDTYSIRNSNHWI---VTTESGREVHARKVIACHILTT 291
Db 232 RVVEALVKIGREBGVDFRMTAVKILLSEDDGVGAKVELEDGRLEADVVVWNSDLVYA 291
QY 292 LDLLGNGGDFRTTLDHWRKIRVGVGIGAVLRATSLATSPSYRGSDATTRESTSGIQLLVSD 351
Db 292 YEKI-----LPKTPVAESILKGRPGSCSSISFYWALDRQ 325
QY 352 ZAHLETAHGAALAGE-----LPRPRAVLGMSFGIDPTTIAPAGRHQVTLW-- 396
Db 326 VPELE-AHNIFLADEYRESFDSIFKGLHLPDEPSFYVNPVSRVDSTAAPGKDSVVVLVP 384
QY 397 -----SQWQFYRLS-----GH-----RWASVABAABADRVIGEME 426
Db 395 VGHLEEDRHASQAHLASRNGHISASPPDQPLTTEKQDWPAMI SLARKTILSTIQ 444
QY 427 APA-PGFTDSVLDRFIOTPRDIESELGMIGNVHVMESLDQMKMLRPLPELSHRVPGA 485
Db 445 SRTNVDLTPLIHSTNSPLSWKQTFNLDRGAILGLSHSPFNVLCFRPTTRA---RKPGA 501
QY 486 -----DGLYLTGASTHPGGGV 501
Db 502 FDAQLLKFGVLGRAAEVLIIDAFRCGRGDKIKGLNMGASAPHTGV 546

RESULT 7
CRTI_NEUCR
ID CRTI_NEUCR STANDARD; PRT; 595 AA.
AC P21334;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase) (Albino-1
protein).
GN AL-1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A / FGSC 987;
RX MEDLINE=90377195; PubMed=21144609;
RA Schmidhauser T.J., Lauter P.R., Russo V.E.A., Vanofsky C.;
RT "Cloning, sequence, and photoregulation of al-1, a carotenoid
biosynthetic gene of Neurospora crassa.";
RL Mol. Cell. Biol. 10:5064-5070(1990).
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
CC -!- COPACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- INDUCTION: By photoinduction.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; M57465; AAA33555.1; -.
PIR; A35919; A35919.
InterPro; IPR002937; Amino oxidase.
InterPro; IPR008150; Bac_phytoene_dh.
InterPro; IPR002005; NAD_BS.
InterPro; IPR008151; Phyt_n dehydro.
PFam; PF01593; Amino_oxidase; 1.

```







carotenoid synthesis by blue light.;

EMBO J. 12:1265-1275(1993).

!- FUNCTION: This enzyme converts phytoene into lycopene via the intermediates of phytofluene, zeta-carotene and neurosporene by the introduction of four double bonds (By similarity).

!- COFACTOR: FAD (Probable).

!- PATHWAY: Carotenoid biosynthesis.

!- INDUCTION: By blue light.

!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M94727; AAA25390.1; -.

PIR; S35306; S35306.

InterPro; IPR000759; Adnck\_reductase.

InterPro; IPR002937; Amino oxidase.

InterPro; IPR008150; Bac\_phytoene\_dh.

InterPro; IPR000205; NAD\_Bs.

InterPro; IPR008151; Phyt\_dehydro.

Pham; PF01593; Amino oxidase; 1.

PRINTS; PR00419; ADKRDFTASE.

PRODOM; PD139017; Phyt\_dehydro; 1.

PROSITE; PS00982; PHYTOENE DH; 1.

Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.

NP BIND 12 45 FAD (ADP PART) (POTENTIAL).

T SEQUENCE 529 AA; 58420 MW; 53536A8DFDD024BC CRC64;

Query Match 6.8%; Score 189.5; DB 1; Length 529;

Best Local Similarity 23.7%; Pred. No. 1.7e-05;

Matches 132; Conservative 74; Mismatches 256; Indels 95; Gaps 25;

8 VVCGSHNALVSAAYLAREGWSVEVLEKQVVLGAVS--TVRFPGYKVDGRSSAHLMI 65

12 IVGAGGGLSALNLAGQGFRTVVRKQAVPGRMKGLTLGASGYAVDTGS---IIQ 68

66 HSGIIEEL-GLGAHGLR-YIDCDPFAFPAPGDTGPGIVFHRDLDTATCOSTIERAGTKD 123

69 LPGVLEQIFRRAARLEDDYVKLLPLDWNTRVHFWDGTHLDTTRHLDRMEALAK-FGPRQ 127

124 ADAYRRFVAVWSERSRVMTAPSTPPGSLNIGAPGLATA-----RGNSELSROFLAPG 178

128 ASALRQWEDGEEKYGIAYQKFC--TSADNLGYAPWRAPLRLPKPMQTLYRQ----- 180

179 DALDDEYFDSEALKAALAWFGAQSGPPMSBPGTAPMVGFAALMHVLP-----PGRVAG 231

181 --LDGFFHDDRVTYALAY-----PSKYLGLHPTT-CSSVSPVIFPLELAFGVWHVSG 229

232 GSGALSALASRAVD-GATVALGDGVTISIRNSNHWTTTSGSRVHARKVIAGCHIIT 290

230 GFRELSRGM-MRCARDIGATFRMGTPEVKR-----VDAGRAVGVK--LVGGEVLD 277

291 TLDLGGGDFDRTLDHWRKIRVPGIGAVLRATLATSALPSVRGDATTRESTSGIQLLVS 350

278 ADVVNADLAYAARSILPAEAREGS-----RLTDAALERAKYSCSTFWAYGLDTVYA 331

351 DRAH-----LRTAHGAALAG-----ELPP-----RPAVLGMSFGSDIPTAPAGRHQV 393

332 DLPHLLIYLSERARTDRDALEDHRVDLEDPFFVVCNPGV-----TDPGAPAGES--- 382

394 TLNSQMPYRLSGHRDWSVAEAEADRIVCMEAFAPGFTDVLDRFIQTFRDIESELG 452

383 TLIVLVPTPTGRVDVWKTQALREKIPAMLEKVLGKGVREHIRERYFTATWEDDEN 442

453 MIGGNVHVEVMSLDQXMLWRPLSELSGHRVPGAD--GLYLTGATPGCGGVSG--ASGRS 508

443 VFRGAVFNLSHTWLQGLPLRP-----KVKNRDIEGLYFVGGGTTPGSGILLTIMESANI 495

QY 509 AARIALSDSRGRKASOW 525

DB 496 AADYLTRAGKGPLPGW 512

RESULT 11

ID CRTI ERWHE STANDARD; PRT; 492 AA.

AC P22871;

DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).

GN CRTI.

OS Erwinia herbicola.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Pantoea.

OX NCBI\_TaxID=549;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EH010;

RX MEDLINE=91088634; PubMed=2263648;

RA Armstrong G.A., Alberti M., Hearst J.E.;

RT "Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes.";

RT Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979(1990).

RL !- FUNCTION: This enzyme converts phytoene into lycopene via the intermediates of phytofluene, zeta-carotene and neurosporene by the introduction of four double bonds.

CC !- COFACTOR: FAD (Probable).

CC !- PATHWAY: Carotenoid biosynthesis.

CC !- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M38423; AAA24820.1; -.

DR EMBL; M87280; AAA64981.1; -.

DR PIR; A39273; A33120.

DR InterPro; IPR002937; Amino oxidase.

DR InterPro; IPR008150; Bac\_phytoene\_dh.

DR InterPro; IPR008151; Phyt\_dehydro.

DR Pfam; PF01593; Amino oxidase; 1.

DR PRODOM; PD139017; Phyt\_dehydro; 1.

DR PROSITE; PS00982; PHYTOENE DH; 1.

KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.

PT NP BIND 5 38 FAD (ADP PART) (POTENTIAL).

SQ SEQUENCE 492 AA; 54503 MW; 8EDC5DB1562083F2 CRC64;

Query Match 6.5%; Score 179.5; DB 1; Length 492;

Best Local Similarity 22.8%; Pred. No. 7.2e-05;

Matches 129; Conservative 62; Mismatches 239; Indels 135; Gaps 24;

QY 8 VVCGSHNALVSAAYLAREGWSVEVLEKQVVLGAVS--TVRFPGYKVDGRSSAHLMI 67

DB 5 VVIGAGGGLALAIRQAAGIPTVLEQRDKPCR-AYVWHQGTFDAGPT---VITDP 60

QY 68 GIIEELGLGAHGLRYIDCDPFAFPAPGDTGPGIVFHR-----DLDTATCOSTIERAC- 119

DB 61 TALEALFTLA-GRMEDY--VRLLPVKP-----FYRLCWESGKTLDYANDSALEAQ 109

QY 120 ----GTCDADAYRRFVAVWSERSRVHMKAFSTPTGTSNLIAGFPGGLATARGNSEL 175

DB 110 ITQPNRDPVGRFRFLAY-----SQAVFQ-----EGYRLGSPVFLS-----PRDML 151

QY 176 APGDAL-----DEYFDSALKAALAWFGAQSGPPMSBPGTAPMVGFAALMHV 223

DB 152 RAGFQLKLQAWQSVQSVRSRFTEDHLRAQAFSHLLVG---GNPFTTSSI---YTLIHA 206

224 LPPGRAV-----GGGSAALASRMAVDGATVALDGVTSIRNSNHW- VTTESGREVH 278  
 207 LERENGWFFPEGCTGALNGMKLFTDLGGIELMARVEELVADNRVSQVRLADGRIFD 266  
 279 ARKVZAGCHILTLDLGNGGDFRITLDHMRKIRVGPICGAVLRLATSALPSVRCDAIT 338  
 267 TDVASNADVNTYKL-----LGHHPVGQKRAALE 298  
 339 RESTSGLOLV---SDRAHLRTAH-----GAALAGELF---PREAVL 374  
 299 RKSNSNLFVLYFLNGQPSLAHTTCFGRYRELIDEITFGSALADDFSLVLSHPCV- 357  
 375 GMSFGIDPTIAPAGHCVTLNSQWQVRLSGHSDWASVAEADRVIGEM-EAFAPGFT 433  
 358 -----TQPSLAPPCASFYIAP-VPHLGNAPLDWAQEGPKLDRIFDYLEERYMPLGR 410  
 434 DSVLDRFIQTPRDIESELGMGMVHMSLDQMLRPLPELSGHRVPGADGLYLTGA 493  
 411 SOLVTQRIFFPADFDHTDLAHLGSAFSLEPLLTQSANFRP-----HNEDSDIANLYLNGA 465  
 494 STHPGGVSG--ASGRSAARTALSD 516  
 466 GTHPGAGIPGVVASAKATASLMIED 490

## SULT 12

TD\_RHOCA  
 CRTD\_RHOCA STANDARD; PRT; 494 AA.  
 17059;  
 01-AUG-1990 (Rel. 15, Created)  
 01-AUG-1990 (Rel. 15, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 Methoxyneurosporene dehydrogenase (EC 1.14.99.-).  
 CRTD.  
 Rhodobacter capsulatus (Rhodospseudomonas capsulata).  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
 Rhodobacteraceae; Rhodobacter.  
 NCBI\_TaxID=1061;  
 [1]  
 SEQUENCE FROM N.A.  
 STRAIN=SB1003 / St. Louis, and BEC404;  
 MEDLINE=89313663; PubMed=2747617;  
 Armstrong G.A., Alberti M., Leach F., Hearst J.E.;  
 "Nucleotide sequence, organization, and nature of the protein  
 products of the carotenoid biosynthesis gene cluster of Rhodobacter  
 capsulatus.";  
 Mol. Gen. Genet. 216:254-268 (1987).  
 [2]  
 SIMILARITY TO CAROTENOID DESATURASES.  
 MEDLINE=90368827; PubMed=2144293;  
 Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;  
 "Carotenoid desaturases from Rhodobacter capsulatus and Neurospora  
 crassa are structurally and functionally conserved and contain  
 domains homologous to flavoprotein disulfide oxidoreductases.";  
 J. Biol. Chem. 265:16020-16024 (1990).  
 -!- FUNCTION: CONVERTS HYDROXYNEUROSPORENE TO DEMETHYLSPHEROIDENE OR  
 METHOXYNEUROSPORENE TO SPHEROIDENE.  
 -!- COFACTOR: FAD (Probable).  
 -!- PATHWAY: Carotenoid and chlorophyll biosynthesis.  
 -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X52291; CAA36537.1; --  
 EMBL; J21165; CAA75544.1; --  
 PIR; S04406; S04406.

DR InterPro; IPR002937; Amino oxidase.  
 DR InterPro; IPR008150; Bac phytoene dh.  
 DR InterPro; IPR008151; Phytin dehydro.  
 DR Pfam; PF01593; Amino oxidase; 1.  
 DR ProDom; PD139017; Phytin dehydro; 1.  
 DR PROSITE; PS00982; PHYTOENE\_DH; 1.  
 KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;  
 KM Oxidoreductase; FAD; Flavoprotein; NAD.  
 FT NP\_BIND 8 41 FAD (ADP PART) (POTENTIAL).  
 SQ SEQUENCE 494 AA; 52312 MW; D1180A023FFEB5A9 CRC64;

Query Match 6.4%; Score 176; DB 1; Length 494;  
 Best Local Similarity 23.4%; Pred. No. 0.00012;  
 Matches 132; Conservative 55; Mismatches 254; Indels 122; Gaps 26;

QY 1 KSAFLDVAVVGSGHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERFPGYKYVDRGSSA 60  
 DB 1 MRSETDVVVGARMGGLAAAGAAAGLVTVVEAGDAPGKARAVT-PGGPADTGTPTV 59  
 QY 61 HLMIRHSGLIEEL-GLGAHGLRYDCDP-----WAFAPAPGTDGPGIVFHRDLDA 110  
 DB 60 -LTMFH--VLDALFAACGTAAEHLTLIPLRLAHFM-----PDGSLDLFTDTEA 108  
 QY 111 TCQSIERACGTKDADAVRRF-----VAVMSRSRHWKAFSTPTGSMILIGAFGLATARG 166  
 DB 109 NIEAIRAFAGDKAAAFRRFDHLTTGLWEATHRSVIAA---PKDLMRIA----- 156  
 QY 167 NSELSRQFLAPG-----DALDEYFDSALKAAALWFGAQSOPMSEPGTAPMVGFPA 219  
 DB 157 -ATVTRPOLWPAIRPGLTMRDLLAHFFDPDLAQLFCRYATYVG---GRPGATP-----A 207  
 QY 220 LMHVLPGRVAVG-----GSGALSAALASRMAVDGATVALDGVTSIRNSNHW- VTT 272  
 DB 208 VLSLIWQAEVQGVWALRECMHGVAAALARVAKGVRFHYGAKAKIRVKEGRVTAIEE 267  
 QY 273 SGREVHARKVIAQCHILFTLD-LIENGSGFDRITLDHWRKIRVGPICGAVLRLATSALPS 331  
 DB 268 TGVSIPTGCACIFNGDPGALRDGLIGDAA--RASME---KSPRPAPSLSAWV-WAFGATP- 320  
 QY 332 YRGDATTRESTSGLOLV---VSDRAHLRTAHGAALAGELPRPAVLGMSFGIDPTIAPA 388  
 DB 321 -----IGVDLAHNVFFTDPELFGPIGAGEMPEEPTLY-----ICAQ 359  
 QY 389 GRHQVTLWSQWQVRL-----SGHRDWSVAEAEADRVIGEMAEAFAPGFTDSVLDRIOT 443  
 DB 360 DREMQAPVPEIERFEIIMNGPAGHPQPPQEEAQCRAITPPMLAAMGLTFSPDPETRALTT 419  
 QY 444 P-----RDIESELGMGMVHMSLDQMLRPLPELSGHRVP---GADGLYLTGAST 495  
 DB 420 PALLSRFPQSLGAIYGG-----SPEGLTATFRPLARTGLKGLYLAGGT 465  
 QY 496 HPGGGVSGA--SGRSAARIALSD 516  
 DB 466 HPGAGVPMALTSHTAARALLAD 488

## RESULT 13

CRTI\_PHYBL  
 ID CRTI\_PHYBL STANDARD; PRT; 583 AA.  
 AC P54982;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).  
 GN CARB.  
 OS Phymyces biakesleanus.  
 OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;  
 OC Phymyces.  
 OX NCBI\_TaxID=4837;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL 1555;  
 RA Ruiz-Hidalgo M.J.;

Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.

-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the intermediary of phytylfuene by the symmetrical introduction of two double bonds at the C-11 and C-11' positions of phytoene.

-!- COFACTOR: FAD (Probable).

-!- PATHWAY: Carotenoid biosynthesis.

-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).

-----

EMBL; X78434; CAAS55:97.1; -.  
PIR; S43139; S43139.  
InterPro; IPR002937; Amino oxidase.  
InterPro; IPR008150; Bac.phytoene\_dh.  
InterPro; IPR000205; NAD\_BS.  
InterPro; IPR008151; Phytin.dehydoro.  
Pfam; PF01593; Amino oxidase; 1.  
ProDom; PD139017; Phytin dehydro; 1.  
PROSITE; PS00982; PHYTOENE\_DH; 1.  
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;  
Transmembrane.  
NP\_BIND 8 41 FAD (ADP PART) (POTENTIAL).  
TRANSMEM 531 551 POTENTIAL.  
SEQUENCE 583 AA; 65983 MW; B0EBF682B12FB591 CRC64;

Query Match 6.4%; Score 176; DB 1; Length 583;  
Best Local Similarity 21.2%; Pred. No. 0.00015;  
Matches 120; Conservative 68; Mismatches 219; Indels 158; Gaps 23;

y 8 VVGGCHNALVSAAVLAREGSHVELEKDTVLGVAVSTVERPPKYVDRCSSAHLMIHRS 67  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
b 8 IIIGAGGTATAARLAEGIKVTIVEXNFGGRCSLINH-NHRDTPQPSLYLMFK-- 64  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
  
y 68 GIIEELG-----LCAGLRVIDCPWAFAPPACTDPGPVFHRDLDATCQS 114  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
b 65 LPEAFPALDEKHEDHVLLRNCHNNYKFED-----DDXKLQLSDDLRRMKE 112  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
  
y 115 IERACGTKDAAYRFVAWSERSHYWKAFSTPTGSNLIGAFCGLATA-RGNSE---- 169  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
b 113 MERLEG---PDGFRLFDMKESHTHYE-----CGVEMAIKQNFEITWK 153  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
  
y 170 LSQQFLAPGDALLD-----EYPDSALKALAWFGAQSG-PMSBPQT----- 211  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
b 154 LIRLQYVALFRHLTFDFVYSRAAKFYFTKKRWMAFTFSQMNGSNPYDSPVNLIQYT 213  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
  
y 212 -----APMWGAALMHVLPGRVVGSGALSALAARMVAVDATVAL-----GDGVT 259  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
b 214 EPAEGIWPYKGSGNTVIQKL-----ENIATEKFGARIYEAPVAKINTDDKKVTVG 265  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
  
y 260 IRNSNHWTVTTESGREVHARKVINGCHIL-----TTLDLLNGGFORTILD-H 307  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
b 266 -----VTLQGSEVTEADAVCWNADLVAYHNLLPPCRWTTNTLAEKLTSSSIISFY 316  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
  
y 308 WRKRIRVPGPGA-VLRIATSALPYRGDATTREBSTSQLQLLVSDRALRTAHGAALAGE 366  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
b 317 WSLK-RVPVELDVHHIFLAEPKSFDFDIETDKMPSELSPVN----- 359  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
  
y 367 LPDRNAVIMGSPSGIDPTIAPAGR-----HQVTLNQSQWOPRYLSGHWDVASVAEA 416  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
b 360 LPSR-----IDTAAPPKGDSMTVLVPIGHMKSKTNAEAEDTYMVKRAEKRVYLE- 408  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
  
y 417 EADRIVGWGEAPAFGFTSVLDRIETQRDTSIELSGMLGMNVHMVENSLDOMQLWRPLPE 476  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
b 409 ----VLERRLGLTNFIDLVERHEFNDSIQVKKNLENMRGSIILGSHDLVQLVFRFSTQ 463  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
  
y 477 LGSHRPVPGADGLYLTGASTHPGGVV 501  
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:

464 DSTGRY---KNLFFVGASTHPTGTV 485

RESULT 14

ID	CRTI_AGRAU	STANDARD;	PRT;	501 AA.
AC	PS4978;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).			
GN	CRTI.			
OS	Agrobacterium aurantiacum.			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.			
NCBI	NCBI_TaxID=44155;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=96062243; PubMed=7592436;			
RA	Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiwara S., Saito T.,			
RA	Ohtani T., Miki M.			
RT	"Structure and functional analysis of a marine bacterial carotenoid			
RT	biosynthesis gene cluster and astaxanthin biosynthetic pathway			
RT	proposed at the gene level."			
RL	J. Bacteriol. 177:6575-6584(1995).			
CC	- - FUNCTION: This enzyme converts phytoene into lycopene via the			
CC	intermediaries of phytofluene, zeta-carotene and neurosporene by			
CC	the introduction of four double bonds (By similarity).			
CC	- - COFACTOR: FAD (Probable).			
CC	- - PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin			
CC	biosynthetic pathway.			
CC	- - SIMILARITY: Belongs to the phytoene dehydrogenase family.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL: D58420; BAA09594.1; "			
DR	InterPro; IPR002937; Amino oxidase.			
DR	InterPro; IPR008150; Bac_phytoene_dh.			
DR	InterPro; IPR008151; Phyt_n_dehydro.			
DR	Pfam; PF01593; Amino oxidase; 1.			
DR	ProDom; PD139017; Phyt_n_dehydro; 1.			
DR	PROSITE; PS00982; PHYTOENE DH; 1.			
KW	Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.			
FT	NP_BIND 12 45 FAD (ADP PART) (POTENTIAL).			
SQ	SEQUENCE 501 AA; 54806 MW; 5F251AF1D679358 CRC64;			
DR	Query Match			
DR	Best Local Similarity 24.1%; Score 174.5; DB 1; Length 501;			
DR	Matches 135; Conservative 51; Mismatches 251; Indels 123; Gaps 24;			
Qy	7 AVVVGSHNALVSNAYLARCGNSVEVLEKDTVLGGAVSTVERPGYKVDKSSAHLMIH 66			
Db	11 AIVGAGFGGLAIARLQSGAGIATTLVEARDKPKGR-AYVHHQGHFLFDAGPT---VITD 66			
Qy	67 SGIIIEGLGANGELRYIDCDPWAFAPAGPTD---GEGIVFHR-----DLDATCQS 114			
Db	67 PDALKEL-----WALTGQDMARDVTLMVPSPFYRLKMPGKGVDFVYVNEADQ 112			
Qy	115 IERAC---GTKDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGLATARGNELS 171			
Db	113 LERQIAQFNPDLEGVRRFRDYAEVYQEGYVKLGTVP-----FLKLGQMLKA----- 160			
Qy	172 RQFLPAGDALLDEVFDEALKAALANFQAQ-----KSHVAKVATFIKDPYLRAQFSYHTLLVGGNPFSTS-----SGPNSEPGTAPMVGF 217			
Db	161 ----APALMKLEAY---KSHVAKVATFIKDPYLRAQFSYHTLLVGGNPFSTS-----SI 207			
Qy	218 AALMHVLPQGRV-----GGSGALSALASAKMAYDGAVALGDGVTISRRNSHWT-VTTE 272			

```

208 YALNHALERGGWFAKGTGNQVAGWALPERLGGOMLNKARVARDTDCPRATGYTLA 267
273 SCREHVARKVAGCHILATL-DLLGNGGFDRTTLDHWRKIRVGGIGAVLRATSLAPS 331
268 DGRALTADWASGVDWENYRDLGH-----TARQSPRAK-----SLNA 306
332 YRGDATTRESTSGQLLVSDRAHLRTAHG-----ALAGELPPRPVGLGMSFS-----GI 381
307 KRWMSLFLVHFLGFEAPKDVAAHTTILFGPRYKELVNEIFKGP-K-LAEDFSLYLHSPCTT 365
382 DPTIAPAG--RHQVTLWQWQPYRLSGHRDWSVAEAEADRIVGEM-EAFAPGFTDSVLD 438
366 DEEMAPPGWSTHYVL--APVPHGLRADIDWAVGPPRYADRLASLSERLIPNLRANLTT 422
439 RIQTPRIESELGIMGNVHVMENSLQMLMRLPLPELSGHRVPGDGLYLRGASTHPG 498
423 TRIFTPDSFASLNAHGSASFVEPILTQSAWFRP-----HNRDKTIRNFVAGTHPG 477
499 GGVSG--ASGRSAAIALSD 516
478 AGIPGVGSAAKATQVMSLD 497
```

## RESULT 15

```

CITI_RHOCA
P17054: STANDARD; PRT; 524 AA.
01-AUG-1990 (Rel. 15, Created)
01-AUG-1990 (Rel. 15, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Phytoene desaturase (EC 1.14.99.-) (Phytoene desaturase).
CITI.
Rhodobacter capsulatus (Rhodospseudomonas capsulata).
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
Rhodobacteraceae; Rhodobacter.
NCBI_TaxID=1061;
[1]
SEQUENCE FROM N.A.
STRAIN=SB1003 / St Louis, and BEC404;
MEDLINE=89313663; PubMed=2747617;
Armstrong G.A., Alberti M., Leach F., Hearst J.E.:
"Nucleotide sequence, organization, and nature of the protein
products of the carotenoid biosynthesis gene cluster of Rhodobacter
capsulatus.";
Mol. Gen. Chem. 216:254-268(1989).
[2]
SEQUENCE FROM N.A.
MEDLINE=89327279; PubMed=2546948;
Bartley G.E., Scolnik P.A.;
"Carotenoid biosynthesis in photosynthetic bacteria. Genetic
characterization of the Rhodobacter capsulatus CrtI protein.";
J. Biol. Chem. 264:13109-13113(1989).
[3]
ERRATUM.
Bartley G.E., Scolnik P.A.;
J. Biol. Chem. 264:18260-18260(1989).
[4]
SIMILARITY TO CAROTENOID DESATURASES.
MEDLINE=90368827; PubMed=2144293;
Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;
"Carotenoid desaturases from Rhodobacter capsulatus and Neurospora
crassa are structurally and functionally conserved and contain
domains homologous to flavoprotein disulfide oxidoreductases.";
J. Biol. Chem. 265:16020-16024(1990).
-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the
intermediary of phytofluene by the symmetrical introduction of two
double bonds at the C-11 and C-11' positions of phytoene.
-!- COFACTOR: FAD (Probable).
-!- PATHWAY: Carotenoid and chlorophyll biosynthesis.
-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

```

CC EMBL; J04969; AA50313.1; -
DR EMBL; X52291; CAA36533.1; -
DR EMBL; Z11165; CAA77540.1; -
DR PIR; A32617; A32617.
DR InterPro; IPR000759; Adnrx reductase.
DR InterPro; IPR002937; Amino Oxidase.
DR InterPro; IPR008150; Bac.phytoene_dh.
DR InterPro; IPR000205; NAD RS.
DR InterPro; IPR008151; Phytin dehydro.
DR Pfam; PF01593; Amino oxidase; 1.
DR PRINTS; PR00419; ADXEDTASE.
DR PROSITE; PS00982; PHYTOENE DH; 1.
DR PROSITE; PD139017; Phytin dehydro; 1.
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
KW Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP BIND 12 45 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 524 AA; 57978 MW; 6425A7E5A06A6B9 CRC64;
```

Query Match 6.3%; Score 174; DB 1; Length 524;

Best Local Similarity 21.7%; Pred. No. 0.00018;

Matches 128; Conservative 76; Mismatches 238; Indels 148; Gaps 28;

```

QY 7 AVVGSCHNALVSAAYLAREGMSVEVLEKDTVLGAVSTVERFFGKYKVDGRSSAHLMIKH 66
DB 11 AVVIGAGLGLAAAMRLGAKGYKTVVDRDLPGRGSSITK-GGHRFDLGT---IVTV 66
QY 67 SGIIIEELGLGHLRYIDC-----DPWAFAPAP-----GTDGPIVPHRDLDTCSIE 116
DB 67 PRLREL-----WADCGRDFDKVSLVWBEFFYTIIDPFGEKYTAGD-DKVKAEV 117
QY 117 RACGTKADAYRFFVAVWSERSHYM-----KAFSTPTGSLNIGAPGLATAR---G 166
DB 118 ARISPGDVEGFRHF--MWDAXARFEYENLGRPKMSKMLDLIKVLPTFGVLRADRSVYG 175
QY 167 NSELRSQFLAPGDALLDEYDFSEALKALAWFGAQSGPMSEPG-----TAPMWGFAALX 221
DB 176 HAK-----KMWKDDHLRFALSPH-----PLFPGDPPHVTSMYLVLSOLE 215
QY 222 HVLPPGRAVGGSGALSAALASRMVADGATVALGDVTSI--RRNSNHWVTVTESGREVHA 279
DB 216 KEFGVHYAIGGVQALADAWAKVITDQGEMLNTEVDEILVSRDGKATGIRLMDGTSLPA 275
QY 280 RKVIAGC---HILPTLLGNGGFDRTTLDHWR-----RKIRVGGIGAVLRATSLAL 329
DB 276 QVWVSNADAGH--TYKRLRN-----RDRWRWTDEKLDKK-----EWSMGLF 315
QY 330 PSYRGDATTRESTSGQLLVSDRAHLRTAHGAA-----LAGELPP-----RPVAV 373
DB 316 VVYFG-----TKGTAKWMDVGHHTVVVGPVRYKHEVQDIFIKGELAEADMSLVVHRPSV 368
QY 374 LGMFSGIDPTIAPAGRHQVTLWQWQPYRLSGHRDWSVAEAEADRIVGEM-EAFAPGF 432
DB 369 -----TDPTAAPKGGDDTFYVLSFVPLNGFDNGVDWSVEAEKYKAKVLRVIERLPGV 421
QY 433 TDSVLDRFIQTPRIESE-LGMIGNVHVMENSLQMLMRLPLPELSGHRV-PCADGLYL 490
DB 422 AEKITEEVVPTPETPRDRLYSLPIGAG-FSLEPRILQSAWFRP-----HNASEVDGLYL 474
QY 491 TGASTHPGGVSGVSGRS-----AARIALSDSRGKASQ 524
DB 475 VGAGTHPGAGVPSVIGSELVAQMIPDAPKPTETPAAAPKARTPRAKAAQ 524
```

Search completed: February 29, 2004, 14:45:20

Job time : 11.9663 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

! protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 14.808 Seconds  
(without alignments)  
3455.835 Million cell updates/sec

tle: US-09-941-947A-38

irect score: 2768

quence: 1 MSAPLDVVVSGHNALVSA.....ALSDSRGKASQWRRSSRS 532

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

tal number of hits satisfying chosen parameters: 283366

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

PIR 78:\*

1: Pirl1:\*

2: Pirl2:\*

3: Pirl3:\*

4: Pirl4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

sult No.	Query Match	Score	Length	DB ID	Description
1	859.5	31.1	511	2 E75561	probable phytoene
2	688	24.9	542	2 S76617	hypothetical prote
3	660.5	23.9	565	2 A12273	hypothetical prote
4	538.5	19.5	543	2 G87635	phytoene dehydroge
5	533	19.3	518	2 G20413	phytoene dehydroge
6	466.5	16.9	544	2 T32588	hypothetical prote
7	439.5	15.9	538	2 B72609	hypothetical prote
8	418	15.1	489	2 F75591	P49 secreted prote
9	340	12.3	472	2 T35936	probable dehydroge
10	327	11.8	469	2 S09189	49K protein - Stre
11	314.5	11.4	536	2 A70523	probable oxidoredu
12	303	10.9	535	2 F70782	probable oxidoredu
13	284	10.3	473	2 C70915	hypothetical prote
14	283	10.2	480	2 A70854	hypothetical prote
15	275	9.9	523	2 T36968	probable phytoene
16	270	9.8	501	2 S75951	hypothetical prote
17	264.5	9.6	517	2 T31463	probable diaphophyt
18	249	9.0	512	2 H84320	phytoene dehydroge
19	247	8.9	536	2 B84327	phytoene dehydroge
20	229.5	8.3	528	2 T51119	phytoene desaturas
21	225	8.1	582	2 T46822	phytoene desaturas
22	224	8.1	495	2 S23633	methoxyneurosporen
23	224	8.1	621	2 T48646	phytoene dehydroge
24	222	8.0	485	2 S49624	methoxyneurosporen
25	214.5	7.7	470	2 T72567	hypothetical prote
26	212.5	7.7	506	2 T34971	probable carotenoi
27	212	7.7	430	2 F90272	phytoene dehydroge
28	211	7.6	380	2 A12195	hypothetical prote
29	208.5	7.5	595	2 A35919	carotenoid biosynt

30	204.5	7.4	518	2 T50745	phytoene dehydroge
31	203.5	7.4	518	2 S49620	phytoene dehydroge
32	202	7.3	485	2 S32171	hydroxyneurosporen
33	200	7.2	511	2 T50910	phytoene dehydroge
34	197.5	7.1	548	2 C75466	phytoene dehydroge
35	191.5	6.9	492	2 D37802	phytoene dehydroge
36	191	6.9	486	2 T50749	methoxyneurosporen
37	189.5	6.8	529	2 S35306	phytoene dehydroge
38	189	6.8	506	2 A82064	hypothetical prote
39	188	6.8	430	2 E84212	hypothetical prote
40	187.5	6.8	492	2 S52586	phytoene dehydroge
41	181	6.5	587	2 A86203	hypothetical prote
42	179.5	6.5	492	2 A33120	phytoene dehydroge
43	176.5	6.4	503	2 AC2446	hypothetical prote
44	176	6.4	494	2 S04406	methoxyneurosporen
45	176	6.4	583	2 S43139	phytoene dehydroge

#### ALIGNMENTS

##### RESULT 1

E75561

probable phytoene dehydrogenase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C:Accession: E75561

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

S: Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: E75561

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-511 <Whi>

A:Cross-references: GB:AE001872; GB:AE000513; NID:G6457750; PIDN:AAF09686.1; PID:G645775

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0093

A:Map position: 1

C:Superfamily: phytoene dehydrogenase

Query Match	31.1%	Score	859.5	DB 2	Length	511			
Best Local Similarity	39.7%	Pred. No.	3.1e-51						
Matches	207	Conservative	80	Mismatches	208	Indels	27	Gaps	11
Qy	6	DAVVVSGHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERFPFGYKVDGSSAHLMI	65						
Db	5	DLIVMGAGHNALVTAAYAAAGLKVGFERRHLVGGAVSTEEVVPVGRFDYGGSAHLIL	64						
Qy	66	HSGTIEBLGAHGLRYIDCDPMAPAPPAGTDPGIVF-HRDLDTACQSI-ERACGTKD	123						
Db	65	MTPVIRELELTRHGLHYLEVDPMFHA-----SDGETPFIHRDAGRTIRELDEKPPG--Q	117						
Qy	124	ADAVRFVAVMSERSHVMKAPSTPTPGSNLIGAFGLATARG-----NSELSRQFLAPG	178						
Db	118	GDVYGRFLDWTTPPARAVADLFNSAPGLDL-----GKQVNRSGQGDWNSQLPR-ILRPY	172						
Qy	179	DALLDEYFDSKAAALAWFCAQSGPMSPEPTAPMVGFPAALMEVLPPGAVGSGSALSA	238						
Db	173	GDVAREVFSERVAPLITWAAQGPFPSPFLSAPFLIMHPLHYEGGVAPKPGSGGLTK	232						
Qy	239	ALASRMAVDGATVALGDGVTSIR-RNSNHMTVTTEGSRVHARKVITAGCHIILTTLDLIGN	297						
Db	233	ALRRATEAGGGEVFTDAPVKVILVKDGAQGIRESGETVTARAVSGVHILTTANAL--	290						
Qy	298	GGFDRITLHDHWRKIRVGPGLVRLATLSALPSYGDATTRETSGLQLLVSDRAHLRT	357						
Db	291	---PAEYVPSAARNVRVNGFGMLLRLLALSEKVKYR-HHTEPDSRIGLGLIINERQIMQ	346						
Qy	358	AHGAALAGELPPRPVAVLGMSPSGIDPTIAPAGRQVTLWSQWQPYRLSGHRDWSVAEAE	417						

```

347 GYGEYLAGQETDPLVAMSFSAVDSDSLAPENGVLWLAQYYPPELA-TGSWEITAE 405
348 ADRIVGEMAFAPQFTOSVLDRTQTQTRDIESELGMLGNNVMEVSLDQMLRPLPEL 477
349 RENTLAFAYAPGTRTIVGELVQTPQLWLGHRGNVMEVSLDQMLRPLPEL 465
350 SHRVPGADGLYTGASTHGGVSGASRAARIALSDSRR 519
351 SQVRWPGVQGLYLTGASTHGGVSGASRAARIALSDSRR 507
352 SQVRWPGVQGLYLTGASTHGGVSGASRAARIALSDSRR 507

RESULT 2
1:6617
;Species: Synechocystis sp. (strain PCC 6803)
;Variety: PCC 6803
;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
;Accession: S76617
;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
;K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
;NA Res. 3, 109-136, 1996
;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
;Reference number: S74322; MUID: 97061201; PMID: 8905231
;Accession: S76617
;Status: preliminary
;Molecule type: DNA
;Residues: 1-542 <KAN>
;Cross-references: EMBL:D46004; GB:AB001339; NID:g1001701; PID:BA010561.1; PID:g100172
;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
;Superfamily: phytoene dehydrogenase

Query Match 24.9%; Score 688; DB 2; Length 542;
Best Local Similarity 31.9%; Pred. No. 1.8e-39;
Matches 175; Conservative 97; Mismatches 229; Indels 48; Gaps 16;

Y 6 DAVVSGSHNALYSAAYLAREGMSVLEKDTVLGAVST----VERPPGYKVDGSSAH 61
b 5 DVVIGAGHNLVCAAYLLQGLVTLLEKREVPVGGAAATEALMPELSPFPNRCADH 64
Y 62 LMRHSGIIEELGLGAGHLYRDCDPAWAPAPAGTDPGIVFHRDLDTQOSIERACGT 121
b 65 EFTLPGVLELGLGAGHLYRDCDPAWAPAPAGTDPGIVFHRDLDTQOSIERACGT 120
Y 122 KDADAYRRFVAVWSERSRHVKAFSTPTG-----SNLIGAPGLTARGNSELS 171
b 121 RDAEKYRQFVYVWIDLVNAVQAFNAPPQALLDLALNYGWNKLSVLAIGSKTKALDFI 180
Y 172 RQFLAPCDALLDEYFDPSEALKALAWFGAQSGPPMSEPTAPMVGFALMHVLPGRVAVG 231
b 181 RTMIGSPEDVLEWFDSEFVKAPLRLCEIGAPPSQKSSSSGMMVAMRHLEGIARPKG 240
Y 232 GSGALSAALASRMVNDGATVALGDGVTSIR-RNSNHWTTTSGREVHARKVIACCHLT 290
b 241 GTGALTALVQLVQAGGKILLDTQTKRVLVNNOAIGVEVANGEQYRAK-----GVIS 295
Y 291 TLD-----LLGNGGDFRTTLDHWR-KIRVGPVIGAVLR--ATSALPSYRGDATTRE 340
b 296 NIDARLFLQLVEPGALAKVQNLGERLRTVNNNEALKIDKALSGLPHTAVAGP-E 354
Y 341 STSGLOLLVSDRAHLRTAAGALAGELPRPVLGMSF-SGIDPTIAPAGRHQVTLMSQ- 398
b 355 DLGTGTLTADSVRHEAALIALGOIPDANPSLYLDIPTVLDPHTAPPGQH--TLWIEF 412
Y 399 WQPYRLSG-----HRDWA-SVAEAEADRIVGEAEAFPGTDSVLDRTTQTRDIESE 450
b 413 PAPYRIAGLEGTGLMGTGWTDSEKLVADRVLDKLTVDYAPNLKSLIIGRVSPAEIAQR 472
Y 451 LGMIGNVWHEVMSLDQMLRPLPELSGHRVPGADGLYLTGASTHGGVSGASRAAR 510
b 473 LGSYNGVYHLDMSLDQMLRPLPELSGHRVPGADGLYLTGASTHGGVSGASRAAR 531

```

```

511 RIALSDSRR 519
512 RIALSDSRR 519
513 RIALSDSRR 519
514 RIALSDSRR 519
515 RIALSDSRR 519
516 RIALSDSRR 519
517 RIALSDSRR 519
518 RIALSDSRR 519
519 RIALSDSRR 519
520 RIALSDSRR 519
521 RIALSDSRR 519
522 RIALSDSRR 519
523 RIALSDSRR 519
524 RIALSDSRR 519
525 RIALSDSRR 519
526 RIALSDSRR 519
527 RIALSDSRR 519
528 RIALSDSRR 519
529 RIALSDSRR 519
530 RIALSDSRR 519
531 RIALSDSRR 519
532 RIALSDSRR 519
533 RIALSDSRR 519
534 RIALSDSRR 519
535 RIALSDSRR 519
536 RIALSDSRR 519
537 RIALSDSRR 519
538 RIALSDSRR 519
539 RIALSDSRR 519
540 RIALSDSRR 519
541 RIALSDSRR 519
542 RIALSDSRR 519
543 RIALSDSRR 519
544 RIALSDSRR 519
545 RIALSDSRR 519
546 RIALSDSRR 519
547 RIALSDSRR 519
548 RIALSDSRR 519
549 RIALSDSRR 519
550 RIALSDSRR 519
551 RIALSDSRR 519
552 RIALSDSRR 519
553 RIALSDSRR 519
554 RIALSDSRR 519
555 RIALSDSRR 519
556 RIALSDSRR 519
557 RIALSDSRR 519
558 RIALSDSRR 519
559 RIALSDSRR 519
560 RIALSDSRR 519
561 RIALSDSRR 519
562 RIALSDSRR 519
563 RIALSDSRR 519
564 RIALSDSRR 519
565 RIALSDSRR 519
566 RIALSDSRR 519
567 RIALSDSRR 519
568 RIALSDSRR 519
569 RIALSDSRR 519
570 RIALSDSRR 519
571 RIALSDSRR 519
572 RIALSDSRR 519
573 RIALSDSRR 519
574 RIALSDSRR 519
575 RIALSDSRR 519
576 RIALSDSRR 519
577 RIALSDSRR 519
578 RIALSDSRR 519
579 RIALSDSRR 519
580 RIALSDSRR 519
581 RIALSDSRR 519
582 RIALSDSRR 519
583 RIALSDSRR 519
584 RIALSDSRR 519
585 RIALSDSRR 519
586 RIALSDSRR 519
587 RIALSDSRR 519
588 RIALSDSRR 519
589 RIALSDSRR 519
590 RIALSDSRR 519
591 RIALSDSRR 519
592 RIALSDSRR 519
593 RIALSDSRR 519
594 RIALSDSRR 519
595 RIALSDSRR 519
596 RIALSDSRR 519
597 RIALSDSRR 519
598 RIALSDSRR 519
599 RIALSDSRR 519
600 RIALSDSRR 519
601 RIALSDSRR 519
602 RIALSDSRR 519
603 RIALSDSRR 519
604 RIALSDSRR 519
605 RIALSDSRR 519
606 RIALSDSRR 519
607 RIALSDSRR 519
608 RIALSDSRR 519
609 RIALSDSRR 519
610 RIALSDSRR 519
611 RIALSDSRR 519
612 RIALSDSRR 519
613 RIALSDSRR 519
614 RIALSDSRR 519
615 RIALSDSRR 519
616 RIALSDSRR 519
617 RIALSDSRR 519
618 RIALSDSRR 519
619 RIALSDSRR 519
620 RIALSDSRR 519
621 RIALSDSRR 519
622 RIALSDSRR 519
623 RIALSDSRR 519
624 RIALSDSRR 519
625 RIALSDSRR 519
626 RIALSDSRR 519
627 RIALSDSRR 519
628 RIALSDSRR 519
629 RIALSDSRR 519
630 RIALSDSRR 519
631 RIALSDSRR 519
632 RIALSDSRR 519
633 RIALSDSRR 519
634 RIALSDSRR 519
635 RIALSDSRR 519
636 RIALSDSRR 519
637 RIALSDSRR 519
638 RIALSDSRR 519
639 RIALSDSRR 519
640 RIALSDSRR 519
641 RIALSDSRR 519
642 RIALSDSRR 519
643 RIALSDSRR 519
644 RIALSDSRR 519
645 RIALSDSRR 519
646 RIALSDSRR 519
647 RIALSDSRR 519
648 RIALSDSRR 519
649 RIALSDSRR 519
650 RIALSDSRR 519
651 RIALSDSRR 519
652 RIALSDSRR 519
653 RIALSDSRR 519
654 RIALSDSRR 519
655 RIALSDSRR 519
656 RIALSDSRR 519
657 RIALSDSRR 519
658 RIALSDSRR 519
659 RIALSDSRR 519
660 RIALSDSRR 519
661 RIALSDSRR 519
662 RIALSDSRR 519
663 RIALSDSRR 519
664 RIALSDSRR 519
665 RIALSDSRR 519
666 RIALSDSRR 519
667 RIALSDSRR 519
668 RIALSDSRR 519
669 RIALSDSRR 519
670 RIALSDSRR 519
671 RIALSDSRR 519
672 RIALSDSRR 519
673 RIALSDSRR 519
674 RIALSDSRR 519
675 RIALSDSRR 519
676 RIALSDSRR 519
677 RIALSDSRR 519
678 RIALSDSRR 519
679 RIALSDSRR 519
680 RIALSDSRR 519
681 RIALSDSRR 519
682 RIALSDSRR 519
683 RIALSDSRR 519
684 RIALSDSRR 519
685 RIALSDSRR 519
686 RIALSDSRR 519
687 RIALSDSRR 519
688 RIALSDSRR 519
689 RIALSDSRR 519
690 RIALSDSRR 519
691 RIALSDSRR 519
692 RIALSDSRR 519
693 RIALSDSRR 519
694 RIALSDSRR 519
695 RIALSDSRR 519
696 RIALSDSRR 519
697 RIALSDSRR 519
698 RIALSDSRR 519
699 RIALSDSRR 519
700 RIALSDSRR 519
701 RIALSDSRR 519
702 RIALSDSRR 519
703 RIALSDSRR 519
704 RIALSDSRR 519
705 RIALSDSRR 519
706 RIALSDSRR 519
707 RIALSDSRR 519
708 RIALSDSRR 519
709 RIALSDSRR 519
710 RIALSDSRR 519
711 RIALSDSRR 519
712 RIALSDSRR 519
713 RIALSDSRR 519
714 RIALSDSRR 519
715 RIALSDSRR 519
716 RIALSDSRR 519
717 RIALSDSRR 519
718 RIALSDSRR 519
719 RIALSDSRR 519
720 RIALSDSRR 519
721 RIALSDSRR 519
722 RIALSDSRR 519
723 RIALSDSRR 519
724 RIALSDSRR 519
725 RIALSDSRR 519
726 RIALSDSRR 519
727 RIALSDSRR 519
728 RIALSDSRR 519
729 RIALSDSRR 519
730 RIALSDSRR 519
731 RIALSDSRR 519
732 RIALSDSRR 519
733 RIALSDSRR 519
734 RIALSDSRR 519
735 RIALSDSRR 519
736 RIALSDSRR 519
737 RIALSDSRR 519
738 RIALSDSRR 519
739 RIALSDSRR 519
740 RIALSDSRR 519
741 RIALSDSRR 519
742 RIALSDSRR 519
743 RIALSDSRR 519
744 RIALSDSRR 519
745 RIALSDSRR 519
746 RIALSDSRR 519
747 RIALSDSRR 519
748 RIALSDSRR 519
749 RIALSDSRR 519
750 RIALSDSRR 519
751 RIALSDSRR 519
752 RIALSDSRR 519
753 RIALSDSRR 519
754 RIALSDSRR 519
755 RIALSDSRR 519
756 RIALSDSRR 519
757 RIALSDSRR 519
758 RIALSDSRR 519
759 RIALSDSRR 519
760 RIALSDSRR 519
761 RIALSDSRR 519
762 RIALSDSRR 519
763 RIALSDSRR 519
764 RIALSDSRR 519
765 RIALSDSRR 519
766 RIALSDSRR 519
767 RIALSDSRR 519
768 RIALSDSRR 519
769 RIALSDSRR 519
770 RIALSDSRR 519
771 RIALSDSRR 519
772 RIALSDSRR 519
773 RIALSDSRR 519
774 RIALSDSRR 519
775 RIALSDSRR 519
776 RIALSDSRR 519
777 RIALSDSRR 519
778 RIALSDSRR 519
779 RIALSDSRR 519
780 RIALSDSRR 519
781 RIALSDSRR 519
782 RIALSDSRR 519
783 RIALSDSRR 519
784 RIALSDSRR 519
785 RIALSDSRR 519
786 RIALSDSRR 519
787 RIALSDSRR 519
788 RIALSDSRR 519
789 RIALSDSRR 519
790 RIALSDSRR 519
791 RIALSDSRR 519
792 RIALSDSRR 519
793 RIALSDSRR 519
794 RIALSDSRR 519
795 RIALSDSRR 519
796 RIALSDSRR 519
797 RIALSDSRR 519
798 RIALSDSRR 519
799 RIALSDSRR 519
800 RIALSDSRR 519
801 RIALSDSRR 519
802 RIALSDSRR 519
803 RIALSDSRR 519
804 RIALSDSRR 519
805 RIALSDSRR 519
806 RIALSDSRR 519
807 RIALSDSRR 519
808 RIALSDSRR 519
809 RIALSDSRR 519
810 RIALSDSRR 519
811 RIALSDSRR 519
812 RIALSDSRR 519
813 RIALSDSRR 519
814 RIALSDSRR 519
815 RIALSDSRR 519
816 RIALSDSRR 519
817 RIALSDSRR 519
818 RIALSDSRR 519
819 RIALSDSRR 519
820 RIALSDSRR 519
821 RIALSDSRR 519
822 RIALSDSRR 519
823 RIALSDSRR 519
824 RIALSDSRR 519
825 RIALSDSRR 519
826 RIALSDSRR 519
827 RIALSDSRR 519
828 RIALSDSRR 519
829 RIALSDSRR 519
830 RIALSDSRR 519
831 RIALSDSRR 519
832 RIALSDSRR 519
833 RIALSDSRR 519
834 RIALSDSRR 519
835 RIALSDSRR 519
836 RIALSDSRR 519
837 RIALSDSRR 519
838 RIALSDSRR 519
839 RIALSDSRR 519
840 RIALSDSRR 519
841 RIALSDSRR 519
842 RIALSDSRR 519
843 RIALSDSRR 519
844 RIALSDSRR 519
845 RIALSDSRR 519
846 RIALSDSRR 519
847 RIALSDSRR 519
848 RIALSDSRR 519
849 RIALSDSRR 519
850 RIALSDSRR 519
851 RIALSDSRR 519
852 RIALSDSRR 519
853 RIALSDSRR 519
854 RIALSDSRR 519
855 RIALSDSRR 519
856 RIALSDSRR 519
857 RIALSDSRR 519
858 RIALSDSRR 519
859 RIALSDSRR 519
860 RIALSDSRR 519
861 RIALSDSRR 519
862 RIALSDSRR 519
863 RIALSDSRR 519
864 RIALSDSRR 519
865 RIALSDSRR 519
866 RIALSDSRR 519
867 RIALSDSRR 519
868 RIALSDSRR 519
869 RIALSDSRR 519
870 RIALSDSRR 519
871 RIALSDSRR 519
872 RIALSDSRR 519
873 RIALSDSRR 519
874 RIALSDSRR 519
875 RIALSDSRR 519
876 RIALSDSRR 519
877 RIALSDSRR 519
878 RIALSDSRR 519
879 RIALSDSRR 519
880 RIALSDSRR 519
881 RIALSDSRR 519
882 RIALSDSRR 519
883 RIALSDSRR 519
884 RIALSDSRR 519
885 RIALSDSRR 519
886 RIALSDSRR 519
887 RIALSDSRR 519
888 RIALSDSRR 519
889 RIALSDSRR 519
890 RIALSDSRR 519
891 RIALSDSRR 519
892 RIALSDSRR 519
893 RIALSDSRR 519
894 RIALSDSRR 519
895 RIALSDSRR 519
896 RIALSDSRR 519
897 RIALSDSRR 519
898 RIALSDSRR 519
899 RIALSDSRR 519
900 RIALSDSRR 519
901 RIALSDSRR 519
902 RIALSDSRR 519
903 RIALSDSRR 519
904 RIALSDSRR 519
905 RIALSDSRR 519
906 RIALSDSRR 519
907 RIALSDSRR 519
908 RIALSDSRR 519
909 RIALSDSRR 519
910 RIALSDSRR 519
911 RIALSDSRR 519
912 RIALSDSRR 519
913 RIALSDSRR 519
914 RIALSDSRR 519
915 RIALSDSRR 519
916 RIALSDSRR 519
917 RIALSDSRR 519
918 RIALSDSRR 519
919 RIALSDSRR 519
920 RIALSDSRR 519
921 RIALSDSRR 519
922 RIALSDSRR 519
923 RIALSDSRR 519
924 RIALSDSRR 519
925 RIALSDSRR 519
926 RIALSDSRR 519
927 RIALSDSRR 519
928 RIALSDSRR 519
929 RIALSDSRR 519
930 RIALSDSRR 519
931 RIALSDSRR 519
932 RIALSDSRR 519
933 RIALSDSRR 519
934 RIALSDSRR 519
935 RIALSDSRR 519
936 RIALSDSRR 519
937 RIALSDSRR 519
938 RIALSDSRR 519
939 RIALSDSRR 519
940 RIALSDSRR 519
941 RIALSDSRR 519
942 RIALSDSRR 519
943 RIALSDSRR 519
944 RIALSDSRR 519
945 RIALSDSRR 519
946 RIALSDSRR 519
947 RIALSDSRR 519
948 RIALSDSRR 519
949 RIALSDSRR 519
950 RIALSDSRR 519
951 RIALSDSRR 519
952 RIALSDSRR 519
953 RIALSDSRR 519
954 RIALSDSRR 519
955 RIALSDSRR 519
956 RIALSDSRR 519
957 RIALSDSRR 519
958 RIALSDSRR 519
959 RIALSDSRR 519
960 RIALSDSRR 519
961 RIALSDSRR 519
962 RIALSDSRR 519
963 RIALSDSRR 519
964 RIALSDSRR 519
965 RIALSDSRR 519
966 RIALSDSRR 519
967 RIALSDSRR 519
968 RIALSDSRR 519
969 RIALSDSRR 519
970 RIALSDSRR 519
971 RIALSDSRR 519
972 RIALSDSRR 519
973 RIALSDSRR 519
974 RIALSDSRR 519
975 RIALSDSRR 519
976 RIALSDSRR 519
977 RIALSDSRR 519
978 RIALSDSRR 519
979 RIALSDSRR 519
980 RIALSDSRR 519
981 RIALSDSRR 519
982 RIALSDSRR 519
983 RIALSDSRR 519
984 RIALSDSRR 519
985 RIALSDSRR 519
986 RIALSDSRR 519
987 RIALSDSRR 519
988 RIALSDSRR 519
989 RIALSDSRR 519
990 RIALSDSRR 519
991 RIALSDSRR 519
992 RIALSDSRR 519
993 RIALSDSRR 519
994 RIALSDSRR 519
995 RIALSDSRR 519
996 RIALSDSRR 519
997 RIALSDSRR 519
998 RIALSDSRR 519
999 RIALSDSRR 519
1000 RIALSDSRR 519

```

RESULT 4

**37635**

ycoene dehydrogenase-related protein [imported] - Caulobacter crescentus  
Species: *Caulobacter crescentus*  
Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
Accession: G87635  
Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
J.; Ermolaeva, M.; White, O.; Salzberg, S.B.; Shapiro, L.; Venter, J.C.; Fraser, C.M;  
. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
Title: Complete Genome Sequence of *Caulobacter crescentus*.  
Reference number: AB87249; MUID:21173698; PMID:11259647  
Accession: G87635  
Status: preliminary  
Molecule type: DNA  
Residues: 1-543 <STO>  
Cross-references: GB:AB005673; NID:g13424781; PIDN:AAK25083.1; GSFPDB:GN00148  
Genetics:  
Gene: CC3121

Query Match            19.5%; Score 538.5; DB 2; Length 543;  
Best Local Similarity     33.1%; Pred. No. 3e+29;  
Matches 180; Conservative       76; Mismatches 234; Indels      53; Gaps     21;

,          6   DAVVVGSGHNALVSAAYLAREGWSVELEKEITVLGGAVSVTVERPPGYKVDRGSASAHLMIR    65  
,          |||||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||::||  
,          8   DAVIDGGHGLVCAPYLAAGLAKLVTCEARGWGGAATFEHPGER-NSVASITYSLL        66  
,          |||||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||  
,          ,          66   HGIIEELGLNHGRLYDDCPWAFAPPAPTGPICVFPHRDDLDATQSIERACCTKDAD       125  
,          :        :        :        :        :        :        :        :        :        :        :        :  
,          67   NPRTIADMCLRELGTFLFERISNFLPI---SDDKYIKLGCGLERLT-QEFRKYSRDAE     122  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          126   AYRRRVFVNMSRSRHWMK--AFSTPPT-GSMILIGAFQGTLATARGNELSQ-----P    174  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          123   VLPAYYAMLDETG-DILRDLAGETPNPLGDGLPLLRAIRQGRLAFLSKRKEDILLDLF    181  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          175   LAPGDALLDYEFDSEALKKAALAANPGAQSPPMSPEFTAPMGVFAALMEVLVP-----    226  
,          :        :        :        :        :        :        :        :        :        :        :        :  
,          182   TKSSARDVLVDGFESPEPVKAAFQ-FDAVVWG-NFASPDT-PGSAYVLLLHTTFCEVNEKKGAW   238  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          227   GRAVGSGALSAAIASRMADVGMTALCDGVTSIRRNSNWT-VTTSSGREVHARKVIAG    285  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          239   GHAVGMGANIQMAKACEAGVEILLDPVEAHIIDGGKAAGVOLVDGQINAPIUSAN       298  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          286   CH-IUTDLLNGNGFD---RTLTDHWRKRIRVPGIGAV-LRLATSALPSYR---GDAT       337  
,          :        :        :        :        :        :        :        :        :        :        :        :  
,          299   VNPELLYKLVPSPALTDFPKADVGYKN-----GSTFFRMNVALSELPSTLCPEGKT    352  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          338   TRETSFGQLLVG----DRAHLRTAKGAALAGELPRPAPVLGKSFSGIDDTIAAGEHQV    393  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          353   AEHHQSGIVAPSILDYMADA-Y-RNAKGGIS----KAPIVELIPSLSDTSLAPPOORHA    407  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          394   TLWSQQPYRISSHGHDMNASAEADRIVGEWEAFAPGFTDSVLDRDIQTTPRIESELGM    453  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          408   SLFCQQAPELPDGSRWDAREAADLIIDTVQWAQFKASVLGRMLSPDLERKFGL        467  
,          |        |        |        |        |        |        |        |        |        |        |        |  
,          454   ICNCVMHVEMSLDQMWRPILPELSHRVPGADGLXYLTGASTHPGGGVSGASGRSAARIA    513  
,          |||||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||  
,          468   TGDDTMHGHSMDQLWATRPLIGHASHRAPTA-GLYNMCAGATHPPGGGVSNPGRNAARI    526  
  
,          514 LSD 516  
,  
,          527 LSD 529

SUULT 5  
0413  
ycoene dehydrogenase related protein [imported] - Sulfolobus solfataricus  
Species: *Sulfolobus solfataricus*  
Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
Accession: G9c413  
She, Q.; Singer, R.K.; Confalonieri, F.; Zivarovic, Y.; Allard, G.; Awayez, M.J.; Chan-  
ong, I.; Jeffries, A.C.; Kozeza, C.J.; Medina, N.; Peng, X.; Thi Ngoc, H.P.; Redder,  
rett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001  
A:Description: Sulfolobus solfataricus complete genome.  
A:Reference number: A39139  
A:Accession: G90413  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-518 <RUT>  
A:Cross-references: GB:A5006641; NID:gl3815726; PIDN:AAK42566.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: SSO2422

Query Match	19.3%	Score 533	DB 2	Length 518
Best Local Similarity	30.3%	Pred. No. 6.8e-29		
Matches 168	Conservative	93	Mismatches 226	Indels 56
Gaps				
QY	5	LDVVVGSHNALVSAAAYLAREGWSVEVLEKOTVLGGAVSTVVERPGYKVDGRGSAHLM	64	
Db	2	IDVAIIIGGHNGLVTAAYLAKAGLKAVAPERRBIVGGASVTEELWPSIKVSTGAVVLSL	61	
QY	65	RHSGIIEBLGAGHLRAVIDCDWAPAPPAGTGDGPIVPHRDLOATQOSIBRACGTGDA	124	
Db	62	RPK-IIEELKLRBGLKVYLDGGLP---LPPENGKKLYIWSLEKTKKEIK-ESKDA	116	
QY	125	DAYRRFVAVM---SERSRHVMKAFSTPP---TGSNLIIGAPGLATARGNSLS----	RQF 174	
Db	117	KNYKKWKFVDFLFAEMADFFM--LNPPEPHIDEASNLINIFRG---NVNBEELALSP	170	
QY	175	LAPZGALIDDEYDFDSEALKAAALA--NFGASGPPMSEPGCTAPMVGPAALMHLVLP	226	
Db	171	MODAKSLDDEYFETDDEVKSALADSVMGTFASE--STPGTA---YVLAHHVIGEVNGVK	224	
QY	227	---GRAVCGSALSAAL---ASRMVDGATVALGDGVTISIRNSHNWTVTTESGREVHAR	280	
Db	225	GAWGYVGGMGGEVTAQALRGAEHLGVSYTNAEYDEV--LVNKRVRGKIKNGXTINAK	282	
QY	281	KVIAGCHLITLTDLLGNGGFDRTTLDHWEKIRVGPICAVLRLA--TSALPSYRGDAT	337	
Db	283	IIVSNADPKTTFPKLIRNAELBEDFL---RRVRALKWGVSFKIVGVLEELDPFGNGKS	338	
QY	338	TRBSTSGIQLIVSDRAHLRPAHGAALAGELPRPAVLGMSFGIDPTTAPAGRHQVTLWS	397	
Db	339	LSPEHIASBLIMPMEYIEKAYDDAALGYSREFPWLSINIQSVDPVAPPQKFSFSG	398	
QY	398	QWQYRLSGHDMASVABEADRIVGEMAFAEGFTDSVLDRIPTPRIDBSELMGNN	457	
Db	399	QVLYVD-SKRD--EMKKIAETTFEIKKEFANFKPIKVE--VLTPLDIERRFGINGEN	453	
QY	458	VWHVEMSLDQMLWRPLPELSCHRVPGADGLYLTGASTHPGGGVSGASGRSAARALADS	517	
Db	454	IFHDMTPDQYFFRPLRGYSDVYTP--IRGLYLCGSGTHPGGGVGTGAPGNARHVKILED	512	
QY	518	RRG 520		
Db	513	RRG 515		

## RESULT 6

T32568  
 Hypothetical protein F37C4.6 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 CDate: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 CAccession: T32568  
 R:Langston, Y.; Sansone, J.; Wohlmann, P.  
 submitted to the EMBL Data Library, December 1997  
 A:Description: The sequence of *C. elegans* cosmid F37C4.  
 A:Reference number: 221193  
 A:Accession: T32568  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-544 <LAN>  
 A:Cross-references: EMBL:AF036705; PFDN:AA95172.1; GSPDB:GM00022; CESP:F37C4.6  
 A:Experimental source: strain Bristol N2; clone F37C4  
 C:Genetics:



Gene: CESP.F37C4.6

Map position: 4

Introns: 29/3; 230/2; 397/2; 484/3

Query Match 16.9%; Score 466.5; DB 2; Length 544;  
Best Local Similarity 28.3%; Pred. No. 2.5e-24;  
Matches 157; Conservative 92; Mismatches 231; Indels 75; Gaps 20;

Y 6 DAVVSGHNAVSAAYLABEGWSEVLEKDTVLGAVSTVERFCYKVDGSSAHLAIR 65  
b 16 DALLIGGHNGLTAAAYLTAKGKVCVLRHVVGGAAVTEIVFGFSPRASLYLLSLR 75  
Y 66 HSGIIEELGAGHLYIDCDPWAFAPPGTDPGCVIFHRDLDTATCQSIERACGTAD 125  
b 76 PVVMQELNKKFGLAYHTRPNPSFTPIR--NTHESLLIGMDMAENQKEIAK-FQORDAG 131  
Y 126 AYRFVAVMSER-----HYMKAFST-PTGSGNLIGAFGLA 162  
b 132 NYPKYHFTSEIVHSFEQLMDYEPDLQPKIHLPLPHLYLLFKTVQLGRNAVDFYELM 191  
Y 163 TARGNSELSQLAPGDALLIDYDFDSEALKAAALAMFG--AQGGPPMSBPGTAPMVGFAL 220  
b 192 T-----APISKINKWFESDLVKATLGTGVGLAASP-DRGT-----GYVLL 234  
Y 221 MHVLP-----GRAVGGSGALSAAASRAVAVGATVALGDGVTISRNSN-HWTVT 271  
b 235 HHVIGGLDEHKGAVYVGGMGAVSNAIAECAKSHGAEIYTEQDQVEVLLDGNVAKGVL 294  
Y 272 ESGREHARVIAAGCHILTTDLGNGGDFRTTLDHWRKIRVGPICIGAV--LRATSA 329  
b 295 SNGKELHSLKVMNSNATPHVFNHLVK--KESLPFEHFNINQIDYTFVTKINAVZEL 351  
Y 330 PSY-----RGDATTRETSGLQLLVSDRAHLRTAHGAAL---AGELPPRAVLGMSFGI 381  
b 352 PNFELAKPNQSGEMPHEQTITEN---NCENMQVVDHDAVMYKNGYSRRPVIEMTIPSSV 408  
Y 382 DPTIA-PAGRHQVTLASQPYRLSCHRWDASVABAE-ADRVGEMAPAGFTDVLDR 439  
b 409 DRTIVSADGHVLLFTQYTPPS-PKDGEMTEKTEYAKGVFSEIDAYAPNFSVIGY 467  
Y 440 FIQTPRDIESELGMIGNVHMEVSLDQMWLRPFLSCHRVPAGDGLYLTGASTHFGG 499  
b 468 DILTPPDQIQTGNTGIFHSGMSLDQVYSRPISKMSNYTP-IESLYLCSGNAHGG 526  
Y 500 GVSGAGSRNARIAL 514  
b 527 GVTGAPCRUSALHAL 541

RESULT 7

Hypothetical protein APE1336 - Aeropyrum pernix (strain K1)

Species: Aeropyrum pernix

Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

Accession: B72609

Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah

wa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

NA, Res. 6, 83-101, 1999

Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr

Reference number: A72450; MUID:99310339; PMID:10382966

Accession: B72609

Status: preliminary

Molecule type: DNA

Residues: 1-538 &lt;KAW&gt;

Cross-references: DDBJ:AF000061; NID:G5104821; PIDN:BAA80328.1; PID:G5105014

Experimental source: strain K1

Gene: APE1336

Superfamily: phytoene dehydrogenase

Query Match 15.9%; Score 439.5; DB 2; Length 538;

Best Local Similarity 29.4%; Pred. No. 1.8e-22;

Matches 172; Conservative 70; Mismatches 227; Indels 117; Gaps 25;

QY 10 VGSCHNALVSAAYLABEGWSEVLEKDTVLGAVSTVERFCYKVDGSSA-----HLMI 64  
DB 1 MGGHNGLTAAASVLAARGLRVALLERERLGLAASYSFPWPGFTAPFGAYVLGLYPRWLM 60  
QY 65 RHSGIIEELGAGHLYIDCDPWAFAPPGT-----DGPGCVIFHRDLDTATCQSIERAC 119  
DB 61 REAGIEGRVRL-----LPKDPGCVTVVLGEGRALRVVGDARRTSRBSRY- 104  
QY 120 GTDADAYRFPVAVMS-----EESRHVMKAFSTPTPGSGNLIGAFGL 161  
DB 105 SPADGGEAYLRWASLWALGALVLLLEVYSHPLSPREVVEHAYRAARTPLVGRIVEAL---- 161  
QY 162 ATARGNSELSQLAPGDALLIDYDFDSEALKAAAL---AWFGAQSGPPMSBPGTAPMVGF 217  
DB 162 -----EAAQWMLTAPARILGENFESWEARAALVEDALVGEAAP--STPGT----GI 208  
QY 218 AALMHVL--PPGR-----AVGSGSALSAAASRAVAVGATVALGDGVTISRNSNHWHT- 268  
DB 209 VLAHHYLGTVSGRGGEWAYVKGGMGLIHALLABAAAEAGASITETGARVSEVIVSGRAVG 268  
QY 269 VTTESGREHARVIAAGCHILTTDLGNGGDFRTTLDH-WRRKIRVGPICIGAVLRLAT 327  
DB 269 VRTWDRVYRARKAVLWAAASIKTLPSV-----VELDRGLARRIRTLLESSGASSKIVLA 321  
QY 328 A-----LPSYR--GDATTRETSGLQLLVSDRAHLRTAHGAALAGELPPRAVLGMSFS 379  
DB 322 AREPLKPAEYRWLGEDLYTSV---ITPMGMEYAEKAYGEAVSRGVSRPEWLSVNVLM 377  
QY 380 GIDPTIAPAGRHQVTLWSO--WQPYRLSCHRWDASVABAEADRVGEMAF-----APGFT 433  
DB 378 RVDPLGAPENGLIASIFLOYTWKPAKSWCEDEKSEVAE-----RGLGVLESVTLPREGVR 433  
QY 434 DSVLDRIOTPRDIESELGMIGNVHMEVSLDQMWLRPFLSCHRVPAGDGLYLTG 493  
DB 434 VEVL-----TPRDYEG-LGNPGGSIPIHISMRDLQLYSSRPLPEASGYTLFGVDRLYLASA 487  
QY 494 STHPGGVSGAGSRNARIALSD-----SRR---GKASQWMMRRSSR 531  
DB 488 SSHPGGVSGELPQWLASRLLEDGLGVEKPRRLSLKALKWARSAR 533

RESULT 8  
F75591  
P49 secreted protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: F75591  
R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, B.K.; Peterson, J.D.; Dodson, R.J.;  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036996; PMID:10567266  
A:Accession: F75591  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-489 <WHI>  
A:Cross-references: GB:AE001863; GB:AB001825; NID:G6460670; PIDN:AAF12427.1; PID:G64607  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRA0363  
A:Map position: 2  
C:Superfamily: phytoene dehydrogenase

Query Match 15.1%; Score 418; DB 2; Length 489;  
Best Local Similarity 30.2%; Pred. No. 4.6e-21;  
Matches 162; Conservative 68; Mismatches 215; Indels 92; Gaps 22;

QY 2 SAEFLDAVVYVSGHNAVSAAYLABEGWSEVLEKDTVLGAVSTVE-RFPQYKVDGSSA 60  
DB 17 SSLLDAVVVGAGPENGGLAAAVTLARAGLRVQVLEAHERVGGGLSSAELTLPGFVHDVGSAL 76



460 HVMSLDQMLWRPLPELSGHRVPGADCLVLTGASTHPGCGVSGASGRSAAR 511  
411 SGAVSGLQLL-RPKISLFPYSTP-HPAVFICSATPPGPGVHGMSGHNAAX 460

## RESULT 11

70523  
probable oxidoreductase - Mycobacterium tuberculosis (strain H37RV)  
;Species: Mycobacterium tuberculosis  
;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
;Accession: A70523  
;Author: S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
;Reference number: A70500; MUID:98295987; PMID:9634230  
;Accession: A70523  
;Status: preliminary; nucleic acid sequence not shown; translation not shown  
;Molecule type: DNA  
;Residues: 1-536 <COL>  
;Cross-references: GB:297188; GB:AL123456; NID:93261805; PIDN:CABL0023.1; PID:e1300070;  
;Experimental source: strain H37RV  
;Genetics:  
;Gene: RV3829C

Query Match 11.4%; Score 314.5; DB 2; Length 536;  
Best Local Similarity 24.3%; Pred. No. 6.2e-14;  
Matches 138; Conservative 73; Mismatches 256; Indels 101; Gaps 18;  
6 DAVVVGSHNALVSAAYLAREGMSVEVLEKDTVLGAVSTVERFPYKVDGSSAHLMI 65  
5 DAIVIGAGHNLTAVALLQRLAGLTACLDKRYAGGMASTVELFDGYPFELAGSVQFP-T 63  
66 HSGIIEELGAGHGLRYIDCDPWAFAPAPCTDGPVGFHRDLDTATQCSIERACGTDAD 125  
64 SSAYSSSELGDS--LPTVDLEVMSVALRGVGD--PVQETDTKMLTHLRVHGA-DRAV 118  
126 AYRFVAVMSRSRHMVKAFSTPTGNSLIGAPGGLATARGNSLSRQFLAPGDALLDEY 185  
119 TGMAGLLAWSQAPTRALGRFEAGTLPKPSDEMAYCATNEPERSAIDDMFLGSVTDVDRH 178  
186 F-DSE--ALKAALAWFAGS--GPPMSEPGTAPWUGFPAALMHLVLPGRV-----CG 232  
179 PPDREKHGALGSMVTLAVNTLYRGP--ATPGSAALAFGL--GVPEGDVFRWKKLRGG 233  
233 SGALSAAASRMAYDGAATVALGDCVTSI-----RRNSNHWVTVTESGREVHARKVIAGCH 287  
234 IGAULTLSQLLERTGGEVRLRSKVTETVDNRSRSARVGVTAAGDTLTSPIVSA-- 291  
288 ILTLDLLGNGGFRTLDHWRKIRVGPVIGAVRLATLSALPSYRGDATTRESTSGLOL 347  
292 -----IAPDVTNELIDPAVLFPSEIRDRYLRIDHRGSYL 325  
348 LVSDRAHLRTAHGAALA-----GELPPRPVAVL 374  
326 ---QMFALQAPFAPAPYQALNDPSMQASMGIFCTPEQVQQQWEDCRRGIVFADPTVV 381  
375 GMSFGIDPTIAPAGREVTLWSQMPYRLSGHRDWSVAEAD---RIVSEMAFAPG 431  
382 LQIPSLHDPSLAPAGKQAASAFAMWFP--IEGSKYGGYGRKAVENGQNVIDKIRLAPN 439  
432 FTDSDVDRFTQTPRDISELGMIKGNVHMVMSLDQMLWRPLPE-LSGHRVPGADGLV 490  
440 FKSGILRYTFTPHMGVMFGAPGDGYCHALLHSDQIGPNRPGKFGIOPPIA-GLYL 498  
491 TGASTHPGCGVSGASGRSAARIALSDSR 518  
499 GSACGCGPGIITPIPGVNAARQALADR 526

## RESULT 12

70782  
probable oxidoreductase - Mycobacterium tuberculosis (strain H37RV)  
;Species: Mycobacterium tuberculosis  
;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
;Accession: F70782  
;Author: S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
;Reference number: A70500; MUID:98295987; PMID:9634230  
;Accession: F70782  
;Status: preliminary; nucleic acid sequence not shown; translation not shown  
;Molecule type: DNA  
;Residues: 1-535 <COL>  
;Cross-references: GB:273101; GB:AL123456; NID:93261565; PIDN:CAA97372.1; PID:gl314036  
;Experimental source: strain H37RV  
;Genetics:  
;Gene: RV0897C  
;Superfamily: phytoene dehydrogenase

Query Match 10.9%; Score 303; DB 2; Length 535;  
Best Local Similarity 24.4%; Pred. No. 3.8e-13;  
Matches 147; Conservative 74; Mismatches 214; Indels 168; Gaps 25;  
6 DAVVVGSHNALVSAAYLAREGMSVEVLEKDTVLGAVSTVERFPYKVDGSSAHLMI-I 64  
9 DVVVVGSHNLVAAAYLAREGLRVLRLERLAQTGGAASVIOAFDGVVEVALSYSLVSL 68  
65 RHSGIIEELGAGHGLRYIDCDPWAFAPAPCTDGPVGFHRDLDTATQCSIERACGTD 123  
69 LPSSIVADLCAPVRLAR----RPFSTYTPAPATAGRSLGILGTPGPRAAHLAAIAGPD 124  
124 ADAYRRF-----VAVW-----SERSRH-----VMKAFSTPTGSSL 154  
125 ABGFAAYRRCRLVTLARLWPTLLEPLREQARDOIVEYGGHEAAAQAQWVDEPIGEAI 184  
155 IGAPGLATARGNSLSRQFLAPGDALLDEYDFSEALKAAALWFGAGSGFPKSEPGTAPM 214  
185 AGAVA-----NDLARGVIAT-DALIGTF-----ARMHEPSLMQN 217  
215 VGPAALMHLVLPGRV-----GGSGALSAALASRMVADGATVALG-----DGVTSIR 261  
218 ICF--LXHLVGGGTGVWHVPIGNGSVTSALATAARHGAEIVTGADVFALDPDG--TVR 273  
262 RNSNHWVTVTESGRE--VHARKVIACGHLITLTDLLGNGGFRTLDHWRKIRVGPVIG 319  
274 YHSD-----GSDGAHLVRGRFVLGVTPAVLASLIG-----EPVAAALAPGAQ 316  
320 AVLELATPSYRGDATT-RESTSGQLIVSDRAHLRTAHGAALAGELPPRPVAVLGMSP 378  
317 VKVMVVRRLPRLRDSVTPQAPAGTFHVNETWSQLDAAISQAASGRLPDPLPCAYCH 376  
379 SGIDPTT-----APAGRHQVTLMSQMPYRLSGHRDWSVAEADRIIVGEMAEFAPGFT 433  
377 SLTDPSILSARLRDAGAQTUTVPGLTPHSPVGD-----TEGLAERLTAVALA---SL 426  
434 DSVLDRIQ-----TPROIESLGMIGGNVHMVMSLDQMLWRPLPEL 477  
427 NSVLAEPIDQVLWTDQSKPCIETTTLDLQRTLTGMTGNIHFHALS-----W----- 474  
478 SGHRVPGADG-----LYLTGASTHPGCGVSGASGRSAARIALS--D 516  
475 -----PPADNDPLDTPARQMGVATDHERIMLCGSGARRGGAAGVGGIGHNAAMAVLACLA 529  
517 SRR 519  
530 SRR 532

## RESULT 13



```

;Genetics:
;Gene: crtB; SCORDB:SCJ1.35
;Superfamily: phytoene dehydrogenase

Query Match      9.9%; Score 275; DB 2; Length 523;
Best Local Similarity 25.9%; Pred No 3e-11;
Matches 150; Conservative 65; Mismatches 223; Indels 142; Gaps 26;

y 8 VVVGSGHNLVSAAYLAREGMSVEVLEKOTVLGAVSTVERPPGYKVDGSSAHLMIKHS 67
b 12 VVVGAGLGLACALHLLGAGRRVTWVERDAGPGGRSGRV-RLGGYELDTGTV-LTMDPHL 69
y 68 GIIEELGLGAGSRYIDCDPWAPAPGDTGPGIVFHRDLATCQSIERAGTKDADAY 127
b 70 ADEFAFVAGSLSLRRLTALDPAACFADGSGALDVHTDGEAMEAEVRFAGPAQAGY 129
y 128 RRFVAVMSER-----SRHVMKAFSTP-----PTGSNL--IGAPGGLATARGNSLSRQF 174
b 130 RD-LRRWLESLYRAQMRRFIDTNFDSPLQLLHPDLARLAALGGFGRLDGRIG-----RF 182
y 175 LAPGDALLDDEYFDSEALKAAALANFQAQSGPPMSEPGTAPMVGFAALMHV-----LPPG 227
b 183 LS--DERLRVVFQALYAGVA-----PARA-LAAYAVIAYMDTVAGVWFFK- 226
y 228 RAVGGSGALSAAALASRMAYDGAATVALGDGVTISIRNSNH-WTVTTESGREVHARKVIAGC 286
b 227 --CGMHALPRAVADAAATAGADLRWSAEVKALERSAGRVRAVHLASGERIACDAVLTC 283
y 287 HILTLDLGNGGFDRTLLDHWRRKIRVGPICGAVLRATSAALPSYRGDATTRESTGLQ 346
b 284 ELSTAYGLLGR-----APRRPARLRHSPSAVILHAG----- 314
y 347 LLVSDR-----AHLRTAHGAA-----LAGELPPRAVLGMSFSGIDPTIAPAGR- 391
b 315 ---TDRTWPHLAHHTLSFGAAWERTPEELTRYGELMSDFSLLITRTTHDPALPPGRL 371
y 392 -----QVTLMSQWOP-YELSGHRDWAQVAEAEADRIVGENEAFAPGFTDS 435
b 372 EYVLAPCPTDVGPDVAWVRDLGPRYR-----ESLVGLESRGLEGFADS 416
y 436 VLDRFICTPRDISSELGCMIGNYMHVEMSLDQMWLRPLPELSGHRVPGADGLYLTGAST 495
b 417 VQELLVTPLDWDAQ-GHAAGSPFSVSHTFQGTGPRP-----RNLYRGLDNVVLACGT 470
y 496 HPGGVSG--ASGR-SAARIALSDSRGKASQWMRSSRS 532
b 471 TPGVGVPTVLVSGKLAARVT-----GGAGSRPARTRRS 504
```

Search completed: February 29, 2004, 14:53:00  
Job time : 17.808 secs



121 HLMIRSGIIEELGLGAGLRYIDCPWAPAPGCTDGPVIFHRDLATQSIERACG 120  
121 TKDADAYRRVAVWSERSRHVWKAFSTPTGSLNIGAFGLATARGNSLSQFLAPGDA 180  
121 TKDADAYRRVAVWSERSRHVWKAFSTPTGSLNIGAFGLATARGNSLSQFLAPGDA 180  
181 LLDYFDSSEALKAAALAWFGAQSGPMPSEPGTAPMWGFAALMHVLPFGRAVGGSGALSAA 240  
181 LLDYFDSSEALKAAALAWFGAQSGPMPSEPGTAPMWGFAALMHVLPFGRAVGGSGALSAA 240  
241 ASRMVADGATVALGDGVTSTIRNSNHWTTTSGREHARKVIAGCHILTTDLGNGGF 300  
241 ASRMVADGATVALGDGVTSTIRNSNHWTTTSGREHARKVIAGCHILTTDLGNGGF 300  
301 DRTTLDHWRKIRVGPICIGAVLRLATSLPSYRGDATTRESTSGQLVLSDRHRLTAHG 360  
301 DRTTLDHWRKIRVGPICIGAVLRLATSLPSYRGDATTRESTSGQLVLSDRHRLTAHG 360  
361 AALAGELPPRPVAVLGMFSFGIDPTIAPAGRHQVTLWSQOPYRLSGHRDWSVAEAD 420  
361 AALAGELPPRPVAVLGMFSFGIDPTIAPAGRHQVTLWSQOPYRLSGHRDWSVAEAD 420  
421 IVGEMAFAPGFTDVLDRPDIOTPRDIESELGMIGGNVHVMESLDMMLWRPLPBLSGH 480  
421 IVGEMAFAPGFTDVLDRPDIOTPRDIESELGMIGGNVHVMESLDMMLWRPLPBLSGH 480  
481 RVPGADGLYTGASTHPGGVSGASGRSAARIALSDSRGKASQWRRSSRS 532  
481 RVPGADGLYTGASTHPGGVSGASGRSAARIALSDSRGKASQWRRSSRS 532

RESULT 2  
US-10-209-372-2  
Sequence 2, Application US/10209372  
Publication No. US20030100045A1  
GENERAL INFORMATION:  
APPLICANT: E. I. du Pont de Nemours, Inc.  
APPLICANT: Cheng, Qiong  
APPLICANT: Tao, Luan  
TITLE OF INVENTION: CAROTENOID KETOLASE GENE  
FILE REFERENCE: CL-1849 US NA  
CURRENT APPLICATION NUMBER: US/10/209,372  
CURRENT FILING DATE: 2002-07-30  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 2  
LENGTH: 532  
TYPE: PRT  
ORGANISM: Rhodococcus erythropolis AN12

US-10-209-372-2  
Query Match 99.9%; Score 2764; DB 14; Length 532;  
Best Local Similarity 99.8%; Pred. No. 8.2e-238;  
Matches 531; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 MSFLDAVVVGGSHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERPPGKYKVDGSSA 60  
1 VSAFLDAVVVGGSHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERPPGKYKVDGSSA 60  
61 HLMIRSGIIEELGLGAGLRYIDCPWAPAPGCTDGPVIFHRDLATQSIERACG 120  
61 HLMIRSGIIEELGLGAGLRYIDCPWAPAPGCTDGPVIFHRDLATQSIERACG 120  
121 TKDADAYRRVAVWSERSRHVWKAFSTPTGSLNIGAFGLATARGNSLSQFLAPGDA 180  
121 TKDADAYRRVAVWSERSRHVWKAFSTPTGSLNIGAFGLATARGNSLSQFLAPGDA 180  
181 LLDYFDSSEALKAAALAWFGAQSGPMPSEPGTAPMWGFAALMHVLPFGRAVGGSGALSAA 240  
181 LLDYFDSSEALKAAALAWFGAQSGPMPSEPGTAPMWGFAALMHVLPFGRAVGGSGALSAA 240

241 ASRMVADGATVALGDGVTSTIRNSNHWTTTSGREHARKVIAGCHILTTDLGNGGF 300  
241 ASRMVADGATVALGDGVTSTIRNSNHWTTTSGREHARKVIAGCHILTTDLGNGGF 300  
301 DRTTLDHWRKIRVGPICIGAVLRLATSLPSYRGDATTRESTSGQLVLSDRHRLTAHG 360  
301 DRTTLDHWRKIRVGPICIGAVLRLATSLPSYRGDATTRESTSGQLVLSDRHRLTAHG 360  
361 AALAGELPPRPVAVLGMFSFGIDPTIAPAGRHQVTLWSQOPYRLSGHRDWSVAEAD 420  
361 AALAGELPPRPVAVLGMFSFGIDPTIAPAGRHQVTLWSQOPYRLSGHRDWSVAEAD 420  
421 IVGEMAFAPGFTDVLDRPDIOTPRDIESELGMIGGNVHVMESLDMMLWRPLPBLSGH 480  
421 IVGEMAFAPGFTDVLDRPDIOTPRDIESELGMIGGNVHVMESLDMMLWRPLPBLSGH 480  
481 RVPGADGLYTGASTHPGGVSGASGRSAARIALSDSRGKASQWRRSSRS 532  
481 RVPGADGLYTGASTHPGGVSGASGRSAARIALSDSRGKASQWRRSSRS 532

RESULT 3  
US-10-209-372-4  
Sequence 4, Application US/10209372  
Publication No. US20030100045A1  
GENERAL INFORMATION:  
APPLICANT: E. I. du Pont de Nemours, Inc.  
APPLICANT: Cheng, Qiong  
APPLICANT: Tao, Luan  
TITLE OF INVENTION: CAROTENOID KETOLASE GENE  
FILE REFERENCE: CL-1849 US NA  
CURRENT APPLICATION NUMBER: US/10/209,372  
CURRENT FILING DATE: 2002-07-30  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 4  
LENGTH: 511  
TYPE: PRT  
ORGANISM: Deinococcus radiodurans R1

US-10-209-372-4  
Query Match 31.1%; Score 859.5; DB 14; Length 511;  
Best Local Similarity 39.7%; Pred. No. 6.8e-68;  
Matches 207; Conservative 80; Mismatches 208; Indels 27; Gaps 11;

6 DAYVGSCHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERPPGKYKVDGSSAHLMLR 65  
5 DLIVMGAGHNALVTAAYARAGLKVGFERRHLVGGAVSTEEVPGYRFDYGGSAHLIR 64  
66 HSGIIEELGLGAGLRYIDCPWAPAPGCTDGPVIFHRDLATQSI-ERACGTXD 123  
65 MTPVIRELELTRHGLHLYLVDPWFHA-----SDGETPWFIRHDAGTIRELDEKPPG--Q 117  
124 ADAYRFAVWWSERSRHVWKAFSTPTGSLNIGAFGLATARG-----NSLSRQFLAPG 178  
118 GDAYGRFLDDTTPFAVADLFPNSAFGLDL-----GKVMRSGQCKDWNEQLPR-ILRPY 172  
179 DALLDYFDSSEALKAAALAWFGAQSGPMPSEPGTAPMWGFAALMHVLPFGRAVGGSGALSAA 238  
173 GDVAREYFSEERVAPLTMWAQSGPPSDPLSAPLLWHPLYHEGGVAPKGGSGGLTK 232  
239 ALASRMVADGATVALGDGVTSTIRNSNHWTTTSGREHARKVIAGCHILTTDLGNG 297  
233 ALARATEAEGGEVFTDAPVKEILVKDQAQIRLESGETTYTARAVVSGVHILTTANAL-- 290  
298 CGFDRITLDHWRKIRVGPICIGAVLRLATSLPSYRGDATTRESTSGQLVLSDRHRL 357  
291 ---PASYFSAARNVAVNGFGMLRLALSEKVKTR-HEFEPDSRIGLGLIKNERQIMQ 346  
358 AHGAALAGELPPRPVAVLGMFSFGIDPTIAPAGRHQVTLWSQOPYRLSGHRDWSVAEAB 417

347 GYGEYLAQPTDPLVAMSFSAVDDSLAPPNGDVLWLAQYYPFELA-TGSEWETRAEA 405  
418 ADRIVGEMAFAPGDTOSVLDRTOTPRDIESELGCMGVNVMVEXSLDOMLWRPLPEL 477  
406 RENILRAFEHVAPGTROTIVGELVOTQWLETNLGLHRGNVHLEKSFDOQFSPFRWLKA 465  
478 SGHRVPGADGLYLTGASTHPGGVSGSGRSAAHIALSDSR 519  
466 SQYRWPGVQGLYLTGASTHPGGGIMGASGRNAARVIVKDLTR 507

## SULT 4

-10-369-493-23397  
Sequence 23397, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 23397  
LENGTH: 511  
TYPE: PRT  
ORGANISM: Deinococcus radiodurans  
-10-369-493-23397

Query Match 31.13; Score 859.5; DB 15; Length 511;  
Best Local Similarity 39.7%; Pred. No. 6.8e-68;  
Matches 207; Conservative 80; Mismatches 208; Indels 27; Gaps 11;

6 DAVVVGSHNALVSAAYLAREGNSVEVLEKDTVLGGAVSTVERFGYKVDGRSSAHLIR 65  
5 DLIWVGHAGNALVTAAYAAAGLKVGFERRHLVGGAVSTEEVVPGRFDYGGSAHLIR 64  
66 HSGIIEELGLGAHGLRYIDCDPWAFAPAPGDTGPGIVF-HRDLATCQSI-ERAQTKD 123  
65 MTPIVRELETRHGLHYLEVDPMFHA-----SDGETPWFIRHDAQTRIRELDEKPG--Q 117  
124 ADAYRRFVAVNSRRHVMKAPSTPTGSLNIGAFGLATARG-----NSELRSQFLAPG 178  
118 GDAYGRFLDDMTPPARAVADLFNSAPGLDL----GQWVRSGQKDMNEQLPR-TLRPY 172  
179 DALDVEYFDSEALKAAALWAFGAQSGPPMSBPGTAPMVGFALMHVLPFGRAVGGSGALSA 238  
173 GDVAREYFSEERVEAPLTWAAQSGPPSPDLSAPFLMHPLVHEGGVAPKGGSGGLTK 232  
239 ALASRMVADGATVALGQVTSIR-RNSNHWVTTEGSRVHARKVIAGCHILTTDLGN 297  
233 ALRERATEEGGEVFTDAPVKEILVXQCAQIRLESGETTARAVVSGVHILTTANAL-- 290  
298 GGFRTTLDHWRKIRVPGGIGAVLRATLATSALPSYRGDATTTRESTSGQLLYSDRAHLMT 357  
291 ---PAEVPSAARNVRVNGNFGMLRLALSEKVKYR-EHTEPDSR:GLGLLNKREQIMQ 346  
358 AHGAALAGEPLPPRAVLGMSFGSDIPTIAPAGRHQVTLMSQWOPYELSGHRDMSVAEAE 417  
347 GYGEYLAQPTDPLVAMSFSAVDDSLAPPNGDVLWLAQYYPFELA-TGSEWETRAEA 405  
418 ADRIVGEMAFAPGDTOSVLDRTOTPRDIESELGCMGVNVMVEXSLDOMLWRPLPEL 477  
406 RENILRAFEHVAPGTROTIVGELVOTQWLETNLGLHRGNVHLEKSFDOQFSPFRWLKA 465  
478 SGHRVPGADGLYLTGASTHPGGVSGSGRSAAHIALSDSR 519

US-10-369-493-2831

Db 466 SQYRWPGVQGLYLTGASTHPGGGIMGASGRNAARVIVKDLTR 507

## RESULT 5

US-10-209-372-6  
Sequence 6, Application US/10209372  
Publication No. US20030100045A1  
GENERAL INFORMATION:  
APPLICANT: E. I. du Pont de Nemours, Inc.  
APPLICANT: Cheng, Qiong  
APPLICANT: Zao, Luan  
TITLE OF INVENTION: CAROTENOID KETOGLASE GENE  
FILE REFERENCE: CL-1849 US NA  
CURRENT APPLICATION NUMBER: US/10/209,372  
CURRENT FILING DATE: 2002-07-30  
PRIOR APPLICATION NUMBER:  
PRIOR FILING DATE:  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 6  
LENGTH: 542  
TYPE: PRT  
ORGANISM: Synecocystis sp. PCC6803  
US-10-209-372-6

Query Match 24.9%; Score 688; DB 14; Length 542;

Best Local Similarity 31.9%; Pred. No. 1.5e-52;  
Matches 175; Conservative 97; Mismatches 229; Indels 48; Gaps 16;

QY 6 DAVVVGSHNALVSAAYLAREGNSVEVLEKDTVLGGAVST-----VERFPQYKVDGRSSAH 61  
Db 5 DWIIVGAGHGLVCAVLLQGLGVLTLEKREVPVGGAAATTEALMPELSQFRFNRCIAIDH 64  
QY 62 LMIRHSGIIEELGLGAHGLRYIDCDPWAFAPAPGDTGPGIVFHRDLATCQSIERACGT 121  
Db 65 EFIFLGPVLQSLNLAQYLEYLFCDPSVF---CHGLDQAFMSYRSLEKTCAMH-ATYSP 120  
QY 122 KDADAYRRFVAVNSRRHVMKAPSTPTG-----SNLIGAFGLATARGNSEL 171  
Db 121 RDAKYRQFVNYWTDLLNAVQPAFAPQALLDLALMYGWNLSKSVLAIAAGSKTALDFI 180  
QY 172 RQFLAPGDALLDEYFDSEALKAAALWAFGAQSGPPMSBPGTAPMVGFALMHVLPFGRAVG 231  
Db 181 RTMTGSPEDVLNWFDSERVKAPLARLCSIGAPPQKSGSSGGMVMAHEHLEGIAPKPG 240  
QY 232 GSGALSAAASRMVADGATVALGQVTSIR-RNSNHWVTTEGSRVHARKVIAGCHILT 290  
Db 241 GTGALTALVKLVQAQGGKILTDQTVKSVLNNQALGVEVANGEQYRACK-----GVIS 295  
QY 291 TLD-----LLNGGDFRFTLDHWR-KIRVGFQIGAVLRL--ATSALPSYRGDATTRE 340  
Db 296 NIDARLFLQLVERGALAKVQNQLGERLERRTVNNNEAILKIDCALSGLPHTAMAGP-E 354  
QY 341 STSGQLVLSDRALRTAHGAALAGELPPRAVLGMSF-SGIDPTIAPAGRHQVTLWSQ- 398  
Db 355 DLTUTILIADSVRHEERHAHALIAGQIPDANPSLYLDIPTVLDPHTWAPGQH--TLWIEF 412  
QY 399 WQPYRLSG-----HRDWA-SVAERAEADRIVGEAFAPGFTSDVLDRIPTPRDIESE 450  
Db 413 PAPYRIAGLEGTGLMGTTDELKEKVDADRVIDKLTDPAPNLKSLIIGRRVSPAEALQ 472  
QY 451 LGMIGGNVHVMVEXSLDOMLWRPLPELSGHRVPGADGLYLTGASTHPGGVSGSGSAA 510  
Db 473 LGSYNGVNYHLDMSLDQMFRLPELPIANYQTP-INKLYLTGAGTHPGGSGISGNGRNCA 531  
QY 511 RIALSDSR 519  
Db 532 RVFLKQQR 540

RESULT 6  
US-10-369-493-2831



Sequence 2831, Application US/10369493  
Publication No. US20030233675A1

## GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 2831

LENGTH: 542

TYPE: PRT

ORGANISM: Synechocystis sp.

US-10-369-493-2831

Query Match 24.9%; Score 688; DB 15; Length 542;

Best Local Similarity 31.9%; Pred. No. 1.5e-52;

Matches 175; Conservative 97; Mismatches 229; Indels 48; Gaps 16;

6 DAVVGSNNALVSAAYLAREGWSVEVLEKDTVLGGAVST-----VERPGKYKDRGSSAH 61

5 DVVVICAGHNGLVCAAYLQRLGVTLLKREVPVGGAAATTEALNPESLPQFRFNCAIDH 64

62 LMRHSGIIEELGAGHGLRYIDCDPWAFAPAPGTDGPGIVFHRDLDTATQSIIRACGT 121

65 EFLFPGVLOELNLAGYGLYLEFCDSVP---CPGLDGAQFMSYSELEKTCABI-ATYSP 120

122 KDADAYRRFVAVMSERSRHVKAFSTPTTG-----SNLIGAFGLATARGNSLS 171

121 RDAEKYRQFVNYWTDLLNAVQPAFNAPPOALLDLALNYCWNELKSVLALAGSKTKALDFI 180

172 RQFLAPGDALLDFYDSEALKALAWFGAQSGPPMSEPCGAPMVGFALMHVLPGRVAVG 231

181 RTWIGSPEDVLEWFWDSERVKAPLARCSEIGAPPSQKSSGMMVAVRHLEGIARPKG 240

232 GSGALSAALASRMVADGATVALGDGVTIS-RNSNHWTVTTSGREVHARKVIACHILT 290

241 GTGALTEALVKLVQAQGGKILTDQTVKRVLVENNOALGVEVANGQYRAKK-----GVIS 295

291 TLD-----LLNGGDFRTLLDHWRR-KIRVGPGLGAVLR--ATSLPSYRGDAITRE 340

296 NIDARLFLOLVBPGLAKVQNQLGERLERTVNNNEALIKIDCALSGLPHTNAGP-E 354

341 STSGLOQLVSDRAHLRTAHGAALAGELPPRPVAVLGMSP-SGIDPTIAPAGRHQVTLWSQ- 398

355 DLGTCTILIADSVRVEEAHALIALGQIPDANPSLYLDIPTVLDPWAPPGQH--TLWIEF 412

399 WQPYRLSG-----HEDWA-SVABAEADRIVGEAEAFAPGTDVSLDRFIOTPRDISE 450

413 FAPYRIAGLEGTGLMGSGWTDELKRVADKVIDKLDYAPNLSLIIGRVSSPAELAQ 472

451 LGMIGGNVHVMESLDOMLWRPLPELSGRVPGADGLYLTGASTHPGGGVSGASGRSAA 510

473 LGSXNGNVYHLDMSLDQWMLRPLPELIANYQTP-INKNLYLTGAGTHPGGSIISGMPGNCA 531

511 RIALSDSR 519

532 RVFLKQQR 540

## RESULT 7

US-10-369-493-18878

Sequence 18878, Application US/10369493

Publication No. US20030233675A1

## GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 18878

LENGTH: 539

TYPE: PRT

ORGANISM: Arabidopsis thaliana

US-10-369-493-18878

Query Match 23.9%; Score 660.5; DB 15; Length 539;

Best Local Similarity 31.7%; Pred. No. 4.2e-50;

Matches 174; Conservative 94; Mismatches 224; Indels 57; Gaps 17;

6 DAVVGSNNALVSAAYLAREGWSVEVLEKDTVLGGAVSTV-----RPGKYKDRGSSAH 61

5 DVVVICAGHNGLVCAAYLQRLGVTLLKREVPVGGAAATTECLPKAPGKFNLCALDH 64

62 LMRHSGIIEELGAGHGLRYIDCDPWAFAPAPGTDGPGIVFHRDLDTATQSIIRACGT 121

65 EFLFPGVLEELNLAGYGLYLEFCDSVP---CPHFDGKYFLAKSLEKTCABIAR-YSE 120

122 KDADAYRRFVAVMSERSRHVKAFSTPTT-----TGSNLIIGAFGLATARGNS-----BLS 171

121 RDAKKAETFEYVQWRAIGAMIPMFNAPPKSIIDIVGNYDITKFDLFSVIGSPNKTLDPI 180

172 RQFLAPGDALLDEYDSEALKALAWFGAQSGPPMSEPCGAPMVGFALMHVLPGRVAVG 231

181 RNMLTSAEDILNWFDFSEFLKAPLARCSEIGAPPSQKTIILGALMMARHNPCHAPRPG 240

232 GSGALSAALASRMVADGATVALGDGVTISIRRNHNTVTTSGREVHARKVIAGCH----- 287

241 GTGALIKALNVLNLSKGGVILTDDQVEK-----VLIDDKKAVGVR-VSGGTEYRAK 290

288 --LTLTLD-----LLNGGDFRTLLDHWRR-KIRVGPGLGAVLR--LATSALPSYRGD 335

291 YGVISNIDAKRLFLQMTDKSDVDADPDLEWELRRIVNNNETILKIDLDLDEPLRFPFH 350

336 ATTRETSGLQLVSDRAHLRTAHGAALAGELPPRP-PAVLGMSFSGIDPTIAPAGRHQVT 394

351 AKDEYLVGSILLIADSVHVEQAHSKCTIGEIPDSDESMVVMPSYLDPTLAPSGKH--T 408

395 LMSQ-WQPYRLSGH-----DNASVABAE-ADRIVGEAEAFAPGTDVSLDRFIOTPR 445

409 VWIEFPAYQIAGAEQTPKGTGMDLKNQVADKVDKLTATYAPNVKTATTARVESP 468

446 DIESELGIGNVHVMESLDOMLWRPLPELSGRVPGADGLYLTGASTHPGGGVSGAS 505

469 ELGERLCAYKNGYHIDMTLDQWFFRPLPELIANYKTP-IDNFLTGTAGTHPGGSIISGMP 527

506 GRSARIAL 514

528 GRNCARVFL 536

## RESULT 8

US-10-369-493-19986

Sequence 19986, Application US/10369493

Publication No. US20030233675A1

## GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 19986  
LENGTH: 537  
TYPE: PRT  
ORGANISM: No. US20030233675altoc punctiforme  
-10-369-493-19986

Query Match 23.5%; Score 651.5; DB 15; Length 537;  
Best Local Similarity 32.0%; Pred. No. 2.7e-49;  
Matches 176; Conservative 90; Mismatches 225; Indels 59; Gaps 17;

6 DAVVVGSHNALYSAAVLAEGMSVEVLKDTVLGAVSTVE----RFGYKVDGSSAH 61

5 DVVLIGHGHLVCAAVLLKAGSVLLKRSVPGGNAITEELPQEAQPKFNLCADH 64

62 LMNHSGIIIEELGAGHGLRYIDCDPWAPAPPAGTDPGIVFHRDLDTQCSIERACGT 121

65 EFHILGPWSELEKYLHVLBCDPVVF---CPHPDGKYLGHKSLEKTCABEIAI- YNE 120

122 KQADYRRFVAVNSERSRHVKAFSTPP-----TGSNLGAFGLATARGNS---ELS 171

121 RDAKYAEFVDYQRAICAMIPFNADPKSIIDIVGNYDIKKFKDLFSVIGSPNKTLDPI 180

172 RQFLAPGDALLDEYFDSALKAAALAWFGAQSPPMSEPGTAPVMVGFPAALMHVLPGRVAG 231

181 RTMLTSAEDLLNEWFDEEFLKAPLARLASELGAPPQSKTLAICAIMAMRHNPQWAPRG 240

232 GSGALSALASRMADGATVALGCVGISIRNSNHWTVTTESGREVHARKVIAG----- 285

241 GTGALVQALVNLVTSKGGVILTQGHVEK-----VLIDDGKAVGVR--VAGGKEYRA 289

286 -CHILITLD-----LLGNGGFPRTTLDHWR--KIRVPGIGAVLR--LATSALEPSYRG 334

290 KYGVISNIDAKSLFLQMTDSDVDGADPDLWERLERIVNNNETILKIDLALDEPLHFFH 349

335 DATTRESTGQLLVSDRAHLRTAHGAALAGELPPR--PAVLGMSFGSIDPTIAPGRHQV 393

350 HAHKDEYLVGSLIADVAHQASHKCTLGEIPDADPSVMYVMPSYLDPITLAPPGKH-- 407

394 TLWSQW-QPYRLSGHR-----DWA-SVAEAEADRVIGEMEAFAFGFTDSVLDRFIQT 444

408 TWLIEYAPQIAGAGTGLKGTWTDLKNKYADRVVDKLDADYAPNVMATIAARVESF 467

445 RDISSELGMIGNNVHMVMSLDQMLWRPLPELSGHRVPGADGLYLTGASTHPGGVSGA 504

468 AELGERLGAVKGYHYVDVTDQIMFFRPLPEIANYKTP-IDNLFITGAGTHFGGSISGM 526

505 SGRSAARIAL 514

527 PGRNARAF 536

RESULT 9  
-10-369-493-12005  
Sequence 12005, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 12005  
; LENGTH: 539  
; TYPE: PRT  
; ORGANISM: Mesorhizobium loti  
US-10-369-493-12005

Query Match 19.4%; Score 537.5; DB 15; Length 539;  
Best Local Similarity 30.4%; Pred. No. 4e-39;  
Matches 168; Conservative 92; Mismatches 218; Indels 75; Gaps 21;

Qy 6 DAVVVGSHNALYSAAVLAEGMSVEVLKDTVLGAVSTVERFPGYKVDGSSAHLMI 65

Db 5 DAVVVGSHNALYSAAVLAEGMSVEVLKDTVLGAVSTVERFPGYKVDGSSAHLMI 64

Qy 66 HSGIIEELGAGHGLRYIDCDPWAPAPPAGTDPGIVFHRDLDTQCSIERACGT 125

Db 65 PE-IMRDLPLPRFGLQVISYEGGAVFT---RDGDVLANRYRDHDAHRRBP- FSRDAE 118

Qy 126 AYRRFVAVNSERSRHVK- AFSPTPTGSNL-----IG-----AFGLATARGNS 173

Db 119 AYRYARDVTQOCRFIQLMRTAPDPTSPKPDIGELLYLGGKFKAGLSAEKALTLRF 178

Qy 174 FLAPGDALLDEYFDSALKAAALAW---FGAQSPPMSEPGTAPVMVGFPAALMHVLP 226

Db 179 TMSISD-FLDEYPTDVIKANFALSIGITAGP--MSPGTA---YVLLHYMGEVDGS 231

Qy 227 ---GRAVGGSALSALASRMADGATVALG---DQVTSIRNSNHWTVTTESGREVHA 279

Db 232 VGWGYARGGMAVTKAASFKASGGTIRTGAEDHVLVSRGKAK--GVVLGGGBEVYG 289

Qy 280 RKTAGACHILTLTLLGNGGFPRTTLDHWR-----KIRVPGIGAVLRATLALPS--- 331

Db 290 KLVSNADVKTFKLVE---EKELPDIPLRRVKNFKIR---GSSGKVNIADLSLPEFA 343

Qy 332 ---YRGDATTRESTGQLLVSDRAHLRTAHGAALAGELPPRPAVLGMSFGSIDPT 384

Db 344 LAXDSPYVRGDMHFTDSIERMERAYDDMK-----AGRWSADPFLDMVIFTLDPT 393

Qy 385 LAPAGRCQVTLWSQPYRLSGHRDNASV-AEAEADRVIGEMEAFAFGFTDSVLDRFIQT 443

Db 394 MAPPKHFMSCFVQYAPPKVNG-RDWTADRDGFAESVVAQIAEYSFGFDRIVHMEVRT 452

Qy 444 PRDIESELGMIGNNVHMVMSLDQMLWRPLPELSGHRVPGADGLYLTGASTHPGGVSG 503

Db 453 PREIEAEVGLTEGNI FQBELTFDQLLFNRPVPGVAYRSP-VGELYMCGSSTHPGGVWG 511

Qy 504 ASGRSAARIALSD 516

Db 512 AGRNAAAEILRD 524

RESULT 10  
US-10-369-493-10809  
; Sequence 10809, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 10809

LENGTH: 521

TYPE: PRT

ORGANISM: Sphingomonas aromaticivorans

S-10-369-493-10809

Query Match 17.5%; Score 484; DB 15; Length 521;

Best Local Similarity 30.5%; Pred. No. 2.3e-34;

Matches 164; Conservative 84; Mismatches 234; Indels 56; Gaps 19;

Y 6 DAVVSGGNALVSAAYLAREGWSVEVLEKDTVLGAVSTVERPFGYKVDGSSAHLMI 65  
b 2 DALIIGGGHGLVCAYLAKAGKVKVLEERDVGGAAVTEEFHFGFNSTASYVSLLR 61  
Y 66 HSGIIEELGLGAGHLYIDCDPWAPAPPGT-----DGPGVFVHRDLATCQSIERACG 120  
b 62 PK-VIADMKLHDYGVRIERTISNPF-PFEDTVKLGGGPG-----RTEAFARFS 110  
Y 121 TKDADAYRFPVAVNSRSRHVMKAPS--TPPTGSLNIGAFGGIAT-----ARGNSE 169  
b 111 KDAESYRYDAA-LEKAVNLVLDISLOTFFNVGGGIAALKAAATQGWPIAKLIDIAIQED 169  
Y 170 LSRQFLAPGDALDEYFDSALKAAALWFAQSG--PPMSEPTAPMVGVFAALMHVLP- 226  
b 170 LLDIPTKSARDFLDGWFEDDHVKSAG-EDAVVGNFAGVSTPGSA-----YVLLHHVFGEV 224  
Y 227 -----GRAVGGSGALSAAALASMAVDGATVALGCVTISIRNSNHWI-VTTESGRVH 278  
b 225 NGKLGAWGSHVGGGGAITQAMAKACVDAGVEISLEAPSVRLVNNKAAAGVKLEGEELY 284  
Y 279 ARKVIAGCHILTTLLGNGGPDRTTLDH-WRRKIR-VGPVGIGAV-LRLATSAIPSYR-- 333  
b 285 APIVAAN-----VGPMLYRQWVDASDLDEDFRRMKNYKTGSTFRMNVALSDELDFKVL 340  
Y 334 -GDATRESTSGQLLVSDRAHLRTHAGALAGELPPRAVLGMSFGIDPTIAPAGRHO 392  
b 341 PGKQLAEHTAGI-ILAPGMDYMDQAFIDAQFGWKKPIVEIKIPSTVDDSLAPPGQEV 399  
Y 393 VTLWSQPVLSCHRDWASVAEAEADRIVGEMAPAPGFTDVLORFTQTPRIESELG 452  
b 400 ASLFCQFAPOLPGRSWDDCREVAELIIDTVNDHAPNFKASVIARIQHSPLDLERKFG 459  
Y 453 MIGNVMVEMSLDQMLWRPLBELSGHRVPGADGLVLTGASTHGGVSGASGRSAA 510  
b 460 LIGGDIHGTWGLDQWAAAPVLNGLGDRSP-IGLYMCGSGTHPGGVTGAPGHNA 516

RESULT 11

JS-10-369-493-20413

Sequence 20413, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 20413

LENGTH: 518

TYPE: PRT

ORGANISM: Rhodospseudomonas palustris

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(518)

OTHER INFORMATION: unsure at all Xaa locations

Query Match 17.0%; Score 471.5; DB 15; Length 517;

Best Local Similarity 29.6%; Pred. No. 2.9e-33;

US-10-369-493-20413

Query Match 17.2%; Score 477; DB 15; Length 518;

Best Local Similarity 30.9%; Pred. No. 9.5e-34;

Matches 170; Conservative 85; Mismatches 216; Indels 80; Gaps 26;

QY 6 DAVVSGGNALVSAAYLAREGWSVEVLEKDTVLGAVSTVERPFGYKVDGSSAHLMI 65  
Db 1 DVLIIGAGNGHGLTCAYLARAGLKVYVRRNVVGGAAVTQEFHFGFR-NSVAATVVSLL 59  
QY 66 HSGIIEELGLGAGHLYIDCDPWAPAPPGTGDGPVIFVHRDLATCQSIERACCTKAD 125  
Db 60 NPKVIADLKLHEHGLRIVERKAKNFL-PAP--DGOYLL--TGSNTTAASLAR-LSAADA 113  
QY 126 AYRPFVA-----VMSKSRH-VMEKAFSTPTGSLNIGAFG-----GIATARGNSELSR 172  
Db 114 AFGFAAELETTIADVLHFLVLA---PP---NLVVQFGLPAIRSSINALETANRLRALTM 167  
QY 173 Q-----FLAPGDALDEYFDSALKAAALWFAQSGPPMS--EPGTAPMVGVFAALMH 222  
Db 168 EQQRLLDLFTCSAGEMLDARFEHDLVK-ALFGDAIVGNVASYAAGSA-----YVMLJHH 222  
QY 223 VLP-----GRAVGGSGALSAAALASMAVDGATVALGCVTISIRNSNHWI-VTTES 273  
Db 223 AFGVNGKGVWGRAIGMGCAITSAVAAAARAGAEIETSAVREVLVEKDRVVGVTLLDD 282  
QY 274 GREVHARKVIAGCH---ILTTL---DLLGNGGPDRTTLDHWRKIRKIVGPGIGAV-LRLAT 326  
Db 283 GENVRAEFVASNVNPKLLYTRLLPQDALPD--VRRRMQNK-----TCSGTFRMNVAL 334  
QY 327 SALPSYR-----GDATRESTSGQLLVSDRAHL-RTAHGAALAGELPPRAVLGMSFG 380  
Db 335 SRLPSFTALPGDGHLTAGIIIAPSLGYMDRAYQDARAGNS-----REPVVEMLIPT 388  
QY 381 IDPTIAPAGRHQVTLNSQMPYRLSGHRDWSVAEAEADRIVGEMAPAPGFTDVLDRP 440  
Db 389 LDUASLAPKQGVASLFCQHVAPLPGASWDDHDEVDLMIAVDVRYAPGFAASVLRQ 448  
QY 441 IOTPRDIESLGMIGNVMVEMSLDQMLWRPLBELSGHRVPGCA-DGLVLTGASTHGG 499  
Db 449 ILSPLDLEREFGLVGGGIFHGALSINOLFAPARELPQADY--FGALKGLYHAGSGAQPN 506  
QY 500 GVSGASGRSAA 510  
Db 507 SVSGAPGHNA 517

RESULT 12

US-10-369-493-4938

Sequence 4938, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 4938

LENGTH: 517

TYPE: PRT

ORGANISM: Burkholderia fungorum

US-10-369-493-4938



b 60 NLSLFAGSAPFRKYANELKTKQGLEPAPVADC-----FASAPP--DGRWFGVNDLEKTASR 113  
Y 115 IERACGTKDADAYRRVAVYKRSRVRKAFSTPTGSGNLIGAFGLATARGNS-----SLS 171  
b 114 M-AAPSAADAATWRKLVAAFPCEAEHLFRLLGSPMSARALACTANWLARKKGVAGALDTG 172  
Y 172 RQFLAPGCDALLDEYFSEALKAAALAWFGAQSGPPMSEPCGTAPWGFALMHLVLP----- 226  
b 173 RLALSSPRTWLETFESPVRATLATWGHLL-----DFAPDIAAGGAVFPFYLESMANQS 225  
Y 227 -GRAVGSAG--LSAALASRMAVDGATVALGDVTSIRNSNHWTVTTESGREVH--ARK 281  
b 226 FGVVLKGGADTIIRALAGNVTSGAGKIUTGAEVSEITVNSKATGCVLTSGETHATKA 285  
Y 282 VIAGCHILITDLDL-----GNGGFDRTDHWRRKIRVCGFGAGVLRATGALPSYRDA 336  
b 286 VIAGVAPKALTGLLPGGSGNAGFDTA-----MKFRVAPGT--MMHLALDELPCWRAGS 339  
Y 337 TTRESSTGILLVSDRAHLRTARGAALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLW 396  
b 340 ELRQ--FAYVELSPSLDANERT--YQAWAGMLPDEPVLVVGQTAIDPSRAPQGRH--VLM 395  
Y 397 SQWQ--PYRLSGH-----RDWASVAEAEADRIVGEAEAFAPGFTSDVLDRIQTPRDI 447  
b 396 VQVRMLPAEITGDAGGKIAPAHWDQVKDAVERVLDIETYPAGLRSKILGRSVFSPIDL 455  
Y 448 ESR-LGMIGNVHVMSLDQMLWRPLPELSCHRVPGADGLYLTCAGSTHPCGGVSGAGS 506  
b 456 ERENPLVGGDQVCGSHHLAQNLFERPARGYAGWNTF-VGNLHLTGAATWPGAGTGAASG 514

## ESULT 15

S-10-369-493-5969

Sequence 5969, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 5969

LENGTH: 544

TYPE: PRT

ORGANISM: Caenorhabditis elegans

S-10-369-493-5969

Query Match 16.9%; Score 466.5; DB 15; Length 544;  
Best Local Similarity 28.3%; Pred. No. 8.8e-33;  
Matches 157; Conservative 92; Mismatches 231; Indels 75; Gaps 20;

Y 6 DAVVVGSGHVALYLAAREGHSVEVLEKDTVLGGAVSTVERFPGYKVDGSSAHLMI 65  
b 16 DALLIGGHNGLTAAYLTKAGKKVCVLERHHVVGAAVTEIIVGFRPSRASYLLSLR 75  
Y 66 HSGIIEELGAGHLRYIDCDPWAPAPPAGTGDGIVFHRDLATQCOSIERACGTDAD 125  
b 76 -PVYMOELNLKFTGLRYIRNPNSFTPIR--NTHSLSLIGMDMAENQKEIAK-PSORDAG 131  
Y 126 AYRREVAVMVMSR-----HYMKAFST-PPTGSNLIGAFGLA 162  
b 132 NYPKYEHFISEIVHSTFQELMDYEPDLQKPIHKLPLHYLLFKTVQPLGLRNADVFLM 191  
Y 163 TARGNELSQFLAPGDALLDEYDSEALKAAALWFG--AQSGPPMSEPGTAPWVGFAAL 220

Db 192 T-----APISKIMNKFESDVLKATLGTGTVIGLAASEM-DPBT-----GYULL 234  
Qy 221 MHVLP-----GRAVGGSGALSAALASRMAVDGATVALGDGVTSTIRNSN-HWTVT 271  
Db 235 HHVIGGLDEHKGAWGVYVCGMAVSNIAEACAKSHGAEIYTEDQVQEVLLDGNVAKGVRL 294  
Qy 272 ESGREYHARKVIAGCHILITDILLGNGGFDRTDLDHWRKIRVCGFGIGAV--LRLATSAL 329  
Db 295 SNGKELHSLKIVMSNATPHVTTFNHLV-----KESLPEEFHNRNINOIDYTSPTVKINVAVKEL 351  
Qy 330 PSY-----RGDAATTRESTGLOLLVSDRAHLRTAHAAL-----AGELPPRPAPVLMGMSFSGI 381  
Db 352 PNLAKPNQSGSEPMPHQTTIEM-----NCENMQVHDAVMDYKNGRYSRFPVIEWTIPSSV 408  
Qy 382 DPTIA-PAGRHQVTLMSQWCPYRLSHRDWASVAEAE-ADRIVGEAEAFAPGFTSDVLD 439  
Db 409 DRTIVDSADGHVVLLFTQYTPFS-PKDGWTEETKTEYAKHVFSIDAYAPNFSSVIGY 467  
Qy 440 FIOTPRDISELQMGIMGNVHVMSLDQMLWRPLPELSCHRVPGADGLYLTCAGSTHPCGG 499  
Db 468 DILTPPDIONTFGITSNIFHGSMSLDQLYVSRPIKWSNYSTP-IESIYLOGSGAHFPG 526  
Qy 500 GVSAGSAGSAAIAL 514  
Db 527 GVTGAPGRLSALHAL 541

Search completed: February 29, 2004, 15:28:09

Job time : 38.3218 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

n on: February 29, 2004, 14:35:44 ; Search time 16.1665 Seconds  
(without alignments)  
1698.885 Million cell updates/sec

tle: US-09-941-947A-38

fect score: 2768

quence: 1 MSAPLDAVVVGSHNALVSA.....ALSDRRGRKQMRSSRS 532

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

tal number of hits satisfying chosen parameters: 389414

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

1: /cgm2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgm2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgm2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgm2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgm2\_6/ptodata/2/iaa/PTCUS.COMB.pep.\*  
6: /cgm2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	225	8.1	582	4	US-09-091-725-17
2	217.5	7.9	494	3	US-08-660-645A-5
3	217.5	7.9	494	3	US-09-298-718-5
4	217.5	7.9	494	3	US-09-546-969-5
5	217.5	7.9	494	3	US-08-980-832-4
6	217.5	7.9	494	4	US-09-547-267-5
7	217.5	7.9	494	4	US-09-920-923B-4
8	200	7.2	414	3	US-09-067-626-4
9	191.5	6.9	492	4	US-07-783-705A-4
10	185	6.7	511	4	US-09-934-903-16
11	165.5	6.0	497	4	US-09-934-903-18
12	146	5.3	610	4	US-09-443-184-55
13	139	5.0	863	4	US-09-252-981A-19574
14	134	4.8	538	4	US-09-252-981A-30706
15	133	4.8	4928	3	US-09-036-987A-5
16	133	4.8	4928	3	US-09-370-700-5
17	133	4.8	4928	4	US-09-603-207-5
18	132	4.8	769	4	US-09-252-991A-17737
19	131.5	4.8	584	4	US-09-252-991A-28032
20	130	4.7	711	4	US-09-252-991A-22259
21	129.5	4.7	481	1	US-08-472-028A-6
22	129.5	4.7	481	2	US-08-808-931-6
23	129.5	4.7	481	3	US-08-808-323-6
24	129.5	4.7	481	3	US-09-050-603A-6
25	129.5	4.7	481	3	US-09-102-420B-6
26	129.5	4.7	481	3	US-09-015-683-6
27	129.5	4.7	481	4	US-09-497-698-6

28 129.5 4.7 483 3 US-09-071-296-6 Sequence 6, Appli  
29 129.5 4.7 483 3 US-09-196-268-6 Sequence 6, Appli  
30 129.5 4.7 483 4 US-09-191-938-6 Sequence 6, Appli  
31 126 4.6 489 1 US-08-095-726-8 Sequence 8, Appli  
32 126 4.6 489 1 US-08-095-726-10 Sequence 10, Appli  
33 126 4.6 489 1 US-08-096-043-8 Sequence 8, Appli  
34 126 4.6 489 1 US-08-096-043-10 Sequence 10, Appli  
35 126 4.6 489 1 US-08-096-623A-8 Sequence 8, Appli  
36 126 4.6 489 1 US-08-096-623A-10 Sequence 10, Appli  
37 124.5 4.5 4302 4 US-09-052-469-8 Sequence 8, Appli  
38 124.5 4.5 4302 4 US-08-422-582-8 Sequence 8, Appli  
39 124.5 4.5 4302 4 US-09-052-262-8 Sequence 8, Appli  
40 124.5 4.5 4339 4 US-09-052-469-6 Sequence 6, Appli  
41 124.5 4.5 4339 4 US-08-422-582-6 Sequence 6, Appli  
42 124.5 4.5 4339 4 US-09-052-262-6 Sequence 6, Appli  
43 123.5 4.5 3562 4 US-09-679-279-14 Sequence 14, Appli  
44 122.5 4.4 528 2 US-08-808-931-10 Sequence 10, Appli  
45 122.5 4.4 528 3 US-08-808-323-10 Sequence 10, Appli

#### ALIGNMENTS

##### RESULT 1

US-09-091-725-17  
; Sequence 17, Application US/09091725  
; Patent No. 6329141  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Improved methods for transforming Phaffia  
; and recombinant DNA for use therein  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster LLP  
; STREET: 2000 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: United States of America  
; ZIP: 20006-1898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/091,725  
; FILING DATE: 23-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 95203620.0  
; FILING DATE: 22-DEC-1995  
; APPLICATION NUMBER: EP 96200943.7  
; FILING DATE: 11-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: E. Victor Donahue  
; REGISTRATION NUMBER: 35,492  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 582 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-091-725-17

Query Match 8.1%; Score 225; DB 4; Length 582;

Best Local Similarity 24.6%; Pred. No. 2e-11;

Matches 137; Conservative 65; Mismatches 228; Indels 126; Gaps 24;

QY 7 ATVCGSHALVSAATLAREGWSVEVELEKTVLGGAVSTVERFPQYKVDGSSAHLM:IR 65

DB 12 AIIIVCGIGGTAARLAREGQVTVFEKNYSGRCSLIER-DGYRFDQGPSLILLPDL 70

QY 66 HSGIIEELGLGHHG-LRYIDCDP---WAFAPPAFTDGPVIFHRDLDATCQSIERACGT 121

db 71 FKQTFDLGKEMDWDLIKCEPNVYCHF-----HDEETFTSTDMALLKREVERFEG- 123  
y 122 KDADAYERFVAVWSESRHVMKAPSTPTGSLNIGAGCGLATARGNELSROF-----L 175  
b 124 --KQDFRFLSFQEAHRYELA-----VVHVLQKNPFGFA-----AFRLQFQIGILAL 171  
y 176 APGDAL---LDEYFDSALKAAALWFGAQSG-PPMSRPGT-----APMVGF 217  
b 172 HPFESIWTRVCRVFKTDRLRVFSFVAVMYMGQSPYSAPGTYSLLQVTELTGEGWYPRGQF 231  
y 218 AALMHVL-----PPRAVGGSGALSAAALASMAVDGATVALGCVTSIRNSNHWTVT 270  
b 232 WQVNTLQIVKRNFSAKFNAPVSQVLLS-PAKDRAT-----GVR 273  
y 271 TESGREYHARKVIAGCHILTTLLDNGGFGDRTTLDHWRKIRVGGIGAVLRLATSALP 330  
b 274 LESGEHHADWIVNADLVYASEHL-----IPDARNK-----IQGLGEVKRSWMA 319  
y 331 SYRGDATTRESTSGQLLVS-DR-AHLRTAKGAALAGE-----LPPRAVLG 375  
b 320 DLVGGKKLKGCSLSLFSYWSMDRIVDGLGHNIFLAEDFKGSGFTIFEELGLPADPSFVY 379  
y 376 MSFSGIDPTTIAPAGRHQVTLWSQWPYRLSGHRDWSVAEADRVGMEAF-----429  
b 380 NVSRIDPSAPEGKDAIVLVP-----CGHIDASN--PDYKLVAPARKFVQTLA 431  
y 430 ----PGTDSVLDRFTQTPRDISELSGMIGNVMHMSLDQMLWRPLPELSGHRVPGA 485  
b 432 KGLPDPFKMIVAEKVHDAPEWKEFNLDKGSILGLAHNFMQVLGFRP-----STHPRY 486  
y 486 DGLYLTCGASTHGGGV 501  
b 487 DKLFFVGASTHPTGTGV 502

## RESULT 2

IS-08-660-645A-5

Sequence 5, Application US/08660645A

Patent No. 6087152

## GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/660,645A

FILING DATE: 07-JUN-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 9510888.9

FILING DATE: 09-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 494 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-660-645A-5

Query Match 7.9%; Score 217.5; DB 3; Length 494;

Best Local Similarity 23.6%; Pred. No. 7.3e-11;

Matches 133; Conservative 61; Mismatches 240; Indels 129; Gaps 25;

Qy 7 AVVGSCHNALVAAYLAREGNSVLEKDTVLGGAVSTVERPPGYKYVDRGSSAHLMIKH 66

Db 4 AIVIGAGFGGLAAILQSNAGIATTIVEARDKPGGR-AVVMNDQGHVFDAGPT---VVVD 59

Qy 67 SGIIIEGLGAGHLYIDCDPWAF-----PPAP-----GTDGEGIVFHRDLDA 110

Db 60 PDSUREL-----HALSQPMERDVTLTPVSFFYRLTWADGRSFEYVNDDE 105

Qy 111 TCQSIERACGTDADAYRRFVAVWSESRH-VMKAFSTP--PTGS--NLIGAPGGIATAR 165

Db 106 LIRQV-ASFPNADVDGTRRPHDYAEVYREGVYKLGCTTFFKLGLGOMLNAAPALMLRLQAYR 164

Qy 166 GNSLSRQFLAPGDALLDDEYFDSSEALKAAALWFGAQSGPPMSEPGTAPWVGFPAALMHVLP 225

Db 165 SVTSMVARFTQ--DPHLROAFSPHTLLV-----GGNPFSTS-----SIYALHALE 208

Qy 226 PGRV-----GGSGALSAALASRMVADGATVALGDVTSIRRSNHWI-VTTESGREVHAR 280

Db 209 RRGVWFPAKGTNQLVAGMVALFERLGGTLLLNARVTRIDTEGDRATGVTLLDGRQLRAD 268

Qy 281 KYIAGCHILTL-DILONGGDFRT-----TLDHRRKIRVGPVIGAVLRLATSALPSYRG 334

Db 269 TVASNGDVHMSYRDLGHTTRGTGRTKAAIILNRQWSMSL-----FVLHFGLSKRPE--- 318

Qy 335 DATTRESTSGQLLVSDRAHLRTAHGAALAGELPPRAVLGMSFSG----- 380

Db 319 -----NLAHSVIFG--PRYKGLVNEIFNGPRLPDDPSMYLHSP 355

Qy 381 --IDPTIAPAG--RHQVTLWSQWPYRLSGHRDWSVAEADRVGMEAF-PGFTDS 435

Db 356 CVTDPSLAPEGMSHYVL---APVPHLGRADVWEAEAFGYAERIFEELEERRAIPDLRKH 412

Qy 436 VLDRFIQTPREDIESELGMIGNVMHMSLDQMLWRPLPELSGHRVPCADGLYLGTAST 495

Db 413 LTVSRIFSPADFTSELSAHGSAPSVPEILTQSARFP-----HNRDRAIPNFYIVGACT 467

Qy 496 HPGGVSG--ASGRSAARIALSD 516

Db 468 HPGAGIPGVVGSAAKATAQVWLSD 490

## RESULT 3

US-09-298-718-5

Sequence 5, Application US/09298718

Patent No. 6124113

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:





335 DATTRETSGLQLVSDRAHLRTAHGAALAGELPPRPVAVLGMSPFG----- 380  
319 -----NLAHSHVIFG--PRYKGLVNEIFNGPRLPDDFSMYLHSP 355  
381 --IDPTIAPAG--RHQVTLWSQMPYRLSGHRDWSVAEADRIVGNEMEAPA-PGFTDS 435  
356 CVTDPSLAPEGSMTHYVL---APVPHLGRADVWEAEAPGYAEIRFEELERRAIPDLRKH 412  
436 VLDRFICTQTRDIESELGMIGGNMVMHSEMSLDQMLWRPLPELSGHRVPGADGLYLTGAST 495  
413 LTWSRIFSPADFTSELSAHHGSFAFSEVPILTQSAWFRP-----HNRDRAIPNFYIVGACT 467  
496 HPGCGVSG--ASGRSAARIALSD 516  
468 HPGAGIPGVVGSAAKATAQVWLSD 490

RESULT 5  
S-08-980-832-4  
Sequence 4, Application US/08980832B  
Patent No. 6291204  
GENERAL INFORMATION:  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tsygankov, Yuri  
TITLE OF INVENTION: Improved Fermentative Carotenoid Production  
FILE REFERENCE: Improved Fermentative Carotenoid  
CURRENT APPLICATION NUMBER: US/08/980,832B  
CURRENT FILING DATE: 1997-12-01  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 4  
LENGTH: 494  
TYPE: PRT  
ORGANISM: Flavobacterium sp. R1534  
S-08-980-832-4

Query Match 7.9%; Score 217.5; DB 3; Length 494;  
Best Local Similarity 23.6%; Pred. No. 7.3e-11;  
Matches 133; Conservative 61; Mismatches 240; Indels 129; Gaps 25;

7 AVVVGSHNALVSAAYLAREGMSVLEKDTVLGGAVSTVERPPGYKVDGRSSAHLMTIRH 66  
4 AVVIGAGFGLALALRLQAGIATTIVEARDKPGGR-AVYVNDQGHVFDAGPT---VVTD 59  
67 SGIIELGLGAGHLRYIDCDPWAPA-----PPAP-----GTGPGIVFHRDLDA 110  
60 PDSLREL-----NALSQPMERDVTLLPVSPFRLTWADGRSFYVNDDE 105  
111 TCOSIERAGCTKDADAYRFBVAVMSERSRH-VMKAFSTP--PTGS--NLIGAFGLATAR 165  
106 LIRQV-ASFNPADVGYRRFHDYAEVREGYKLGTTTFFLKQMLNAPALMELQAYR 164  
166 GNSLSRQFLAPGDALLDEYFDSALKALANFAGSQSGPPSEPGTAPWVGFALMHVLP 225  
165 SVHSMVARIQ--DPHLROAFSFHTLLV-----GGNPFSTS-----SIYALIHAE 208  
226 PGRVAV---GGGSAALASAMVADGATVALGDGVTISIRRNSEFWT-VITESGREVHAR 280  
209 RRGGVWFAGGTTNQLVAGVVALFRLGGTLLNARVTRIDTSGDRATCVLLDGRQLRAD 268  
281 KVTAGCHILTL-DLLCNGGFPRT-----TLDHWRKRVGPGICGAVLRLATLSALPSVRG 334  
269 TVASNGDMVSHSRDLGLHTRGRRTKAAILNRQWMSL-----FVLHPLSKRPE--- 318  
335 DATTRETSGLQLVSDRAHLRTAHGAALAGELPPRPVAVLGMSPFG----- 380  
319 -----NLAHSHVIFG--PRYKGLVNEIFNGPRLPDDFSMYLHSP 355  
381 --IDPTIAPAG--RHQVTLWSQMPYRLSGHRDWSVAEADRIVGNEMEAPA-PGFTDS 435  
356 CVTDPSLAPEGSMTHYVL---APVPHLGRADVWEAEAPGYAEIRFEELERRAIPDLRKH 412  
436 VLDRFICTQTRDIESELGMIGGNMVMHSEMSLDQMLWRPLPELSGHRVPGADGLYLTGAST 495

Db 413 LTWSRIFSPADFTSELSAHHGSFAFSEVPILTQSAWFRP-----HNRDRAIPNFYIVGACT 467  
Qy 496 HPGCGVSG--ASGRSAARIALSD 516  
Db 468 HPGAGIPGVVGSAAKATAQVWLSD 490

RESULT 6  
US-09-547-267-5  
Sequence 5, Application US/09547267  
Patent No. 6613543  
GENERAL INFORMATION:  
APPLICANT: Hohmann, Hans-Peter  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tessier, Michel  
APPLICANT: van Loon, Adolphus  
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: NJ  
COUNTRY: USA  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/09/547,267  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA: 08/660,645  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pokras, Bruce A.  
REGISTRATION NUMBER: 32,748  
REFERENCE/DOCKET NUMBER: RAN 6002/170  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-5801  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-547-267-5

Query Match 7.9%; Score 217.5; DB 4; Length 494;  
Best Local Similarity 23.6%; Pred. No. 7.3e-11;  
Matches 133; Conservative 61; Mismatches 240; Indels 129; Gaps 25;

7 AVVVGSHNALVSAAYLAREGMSVLEKDTVLGGAVSTVERPPGYKVDGRSSAHLMTIRH 66  
4 AVVIGAGFGLALALRLQAGIATTIVEARDKPGGR-AVYVNDQGHVFDAGPT---VVTD 59  
67 SGIIELGLGAGHLRYIDCDPWAPA-----PPAP-----GTGPGIVFHRDLDA 110  
60 PDSLREL-----NALSQPMERDVTLLPVSPFRLTWADGRSFYVNDDE 105  
111 TCOSIERAGCTKDADAYRFBVAVMSERSRH-VMKAFSTP--PTGS--NLIGAFGLATAR 165  
106 LIRQV-ASFNPADVGYRRFHDYAEVREGYKLGTTTFFLKQMLNAPALMELQAYR 164  
166 GNSLSRQFLAPGDALLDEYFDSALKALANFAGSQSGPPSEPGTAPWVGFALMHVLP 225  
165 SVHSMVARIQ--DPHLROAFSFHTLLV-----GGNPFSTS-----SIYALIHAE 208

226 PGRV-----GSGGSAALASRMVADGATVALGCVTTSIRNSNHW-TVTESGREVHAR 280  
209 RRGVWFAKGTNQLVAGWALFERLGOTLLNARVTRIDTEGDRATGVTLDSRQLRAD 268  
281 KVIAGCHIITL-DLLGNGGDR-----TLDHWRKIRVCPGIGAVLRATSAIYSRG 334  
269 TVASNDVMSYRDLGHTRRGRKTAAILNRQMSML-----FVLFGLSKRPE--- 318  
335 DATTRESTSGQLLVSDRAHLRTAHGAALAGELPPRPVAVLGMSPG----- 380  
319 -----NLAHHSVIFG--PRYKGLVNEIFNGRPLPDDPSMYLHSP 355  
381 --IDPTIAPAG--RHQVTLMSQWOPYRLSHRDWASVABEADRIVGEAFA-PGFTDS 435  
356 CVTDPSLAPGWSMTHYVL---APVPHLGRADVDEAEAPGYAERIPEELERRAIPDLRKH 412  
436 VLDRIPTQPRDIESELGMIGNVHVMESLDQWMLRPLPELSGHRVPGADGLYLTGAST 495  
413 LTVSRIFSPADPSTELSAHSGSAFSVEPILTQSAWFRP-----HNRDRAIPNFIIVGAGT 467  
496 HPGGVUSG--ASGRSAARIALSD 516  
468 HPGAGIPGVVGSAKATAQVWLSD 490

## SULT 7

-09-920-923B-4  
Sequence 4, Application US/09920923B  
Patent No. 6677134  
GENERAL INFORMATION:  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tsygankov, Yuri  
TITLE OF INVENTION: Fermentative Carotenoid Production  
FILE REFERENCE: 15464 US (C38435/125944)  
CURRENT APPLICATION NUMBER: US/09/920.923B  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 08/980,832  
PRIOR FILING DATE: 1997-12-01  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 4  
LENGTH: 494  
TYPE: PRT  
ORGANISM: Flavobacterium sp. R1534  
-09-920-923B-4

Query Match 7.9%; Score 217.5; DB 4; Length 494;  
Best Local Similarity 23.6%; Pred. No. 7.3e-11;  
Matches 133; Conservative 61; Mismatches 240; Indels 129; Gaps 25;  
7 AVVYSGHNALVSAAYLAREGWSVEVLEKDTVLGGAVSTVERPFGYKVDGRGSSAHLMIHR 66  
4 AIVIGAGFGGLALAIRLQSAAGIATTIVEARDKPGGR-AVVMNDQGHVFDAGPT---VVTD 59  
67 SGIIIEELGLNAGRLYIDCPWAPA-----PPAP-----GTDGPGIVFHRDLDA 110  
60 PDSUREL-----WALSGQPMERDVTLLFVSPFYRLTWADGRSFEYVNDDE 105  
111 TCQSIERACGTDADAYRFVAVMSERH-VMKAFSTP--PTGS--NLIGAFGGLATAR 165  
106 LIRQV-ASFENPADVDGVRHFDYAEVYRGYKLGTTPTPLKGLQMLNAAPALMRLOAYR 164  
166 GNSLSRQFLAPGALLDEYFDSEALKAAALWFGAGSGPPMSEPGTAPVMVGFALMHVLP 225  
165 SVHSKVARFIQ--DPHLRQAFSFHTLLV-----GGNPFSTS-----SIYALIHAE 208  
226 PGRV-----GSGGSAALASRMVADGATVALGCVTTSIRNSNHW-TVTESGREVHAR 280  
209 RRGVWFAKGTNQLVAGWALFERLGOTLLNARVTRIDTEGDRATGVTLDSRQLRAD 268  
281 KVIAGCHIITL-DLLGNGGDR-----TLDHWRKIRVCPGIGAVLRATSAIYSRG 334

DB 269 TVASNDVMSYRDLGHTRRGRKTAAILNRQMSML-----FVLFGLSKRPE--- 318  
QY 335 DATTRESTSGQLLVSDRAHLRTAHGAALAGELPPRPVAVLGMSPG----- 380  
DB 319 -----NLAHHSVIFG--PRYKGLVNEIFNGRPLPDDPSMYLHSP 355  
QY 381 --IDPTIAPAG--RHQVTLMSQWOPYRLSHRDWASVABEADRIVGEAFA-PGFTDS 435  
DB 356 CVTDPSLAPGWSMTHYVL---APVPHLGRADVDEAEAPGYAERIPEELERRAIPDLRKH 412  
QY 436 VLDRIPTQPRDIESELGMIGNVHVMESLDQWMLRPLPELSGHRVPGADGLYLTGAST 495  
DB 413 LTVSRIFSPADPSTELSAHSGSAFSVEPILTQSAWFRP-----HNRDRAIPNFIIVGAGT 467  
QY 496 HPGGVUSG--ASGRSAARIALSD 516  
DB 468 HPGAGIPGVVGSAKATAQVWLSD 490

RESULT 8  
US-09-067-626-4  
Sequence 4, Application US/09067626  
Patent No. 6177086  
GENERAL INFORMATION:  
APPLICANT: Riley, Lee W.  
APPLICANT: Nathan, Carl F.  
APPLICANT: Ehart, Sabine  
TITLE OF INVENTION: DNA MOLECULE CONFERRING ON MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS RESISTANCE AGAINST ANTIMICROBIAL REACTIVE  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
STREET: Clinton Square, P.O. Box 1051  
CITY: Rochester  
STATE: New York  
COUNTRY: USA  
ZIP: 14603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/067,626  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/045,688  
FILING DATE: 06-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldman, Michael L.  
REGISTRATION NUMBER: 30,727  
REFERENCE/DOCKET NUMBER: 19603/491  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 263-1304  
TELEFAX: (716) 263-1600  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 414 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-067-626-4

Query Match 7.2%; Score 200; DB 3; Length 414;  
Best Local Similarity 24.5%; Pred. No. 2e-09;  
Matches 113; Conservative 57; Mismatches 205; Indels 86; Gaps 19;  
QY 72 ELGAGKGLRYIDCPWAFAPAGTGTGPGIVFHRDLATCQSIERACGTDADATRFV 131  
DB 1 EFDLPARGVT-LTVPTIAYANPLPGR--PAAIYHDLAHTCAKLD-----DGASWRRL 51

Y 132 AYWSERSRHVWKAFTPTGSGNLGAFGLATARGN-SELSEQFLAPGD-----ALLDEYF 186  
b 52 GFLVAHSETWVEF-----MLSDKSLPTALGSLVRLGLRLMAQQTTPAWRSAGE-- 100  
Y 187 DGEALKAALAWFCAQSGPPMSFPGTAPVMVGAALMHVLPVPGRAVGGGALSAALASMAV 246  
b 101 DARALFTGVAHAHISPLPSLVAGAGLML--ATLAHSVGMPIPVGQTQAIADALIADLRA 158  
Y 247 DGATVALGCVTSIRNSNHWTTTSGREVIARVIAACHLITLTDLLGNGGDFRTILD 306  
b 159 HGCRLAAGEVITEPQBSVVVFDTAPALLRVY-----RDKLP 195  
Y 307 H-----WRRKIRVPGGICAVLRATLSAL-----PSYRGDATTRESTGLQLLVSDRAHLRTA 358  
b 196 HRVAKALRYRFRAGIAKVDFVLSDSEIPMSDFPLRRAATLH-----LGSTRDQWABA 247  
Y 359 HGAALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLWSQVPLSGHRDOWASVAEAEA 418  
b 248 EADVAAGREHADPMVLAAC-----PHVADPGRIDET---GRAPFWTYAHVPSGSTLDA-T 298  
Y 419 DRIVGEMBAFAPGFTDSVL-DRFIQTPRDISELGMIGNVMEVMSLDQMLWR-----P 473  
b 299 EIVTSVLERFAPGFRDIVAGRAVPAARMADENAVYVGGDI-----TVGANSTWEAIAGP 353  
Y 474 LPELSGHRVPGADGLYLTCASTHPPGGVSGASGRSAARIAL 514  
b 354 TPLRLNWRTP-IPKVYLCSAATPPGAGVHGMCGWYAARTLL 393

## RESULT 9

S-07-783-705A-4

Sequence 4, Application US/07783705A

Patent No. 5429939

GENERAL INFORMATION:

APPLICANT: Misawa, No. 5429939ihiko

APPLICANT: Kobayashi, Kazuo

APPLICANT: Nakamura, Katsumi

APPLICANT: Yamano, Shigeyuki

TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE

SYNTHESIS OF CAROTENOIDS

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ladas &amp; Parry

STREET: 26 West 61 Street

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10023

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: N/A

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/783,705A

FILING DATE: 19911023

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 1-103078

FILING DATE: 21-APR-1989

APPLICATION NUMBER: JP 2-53225

FILING DATE: 05-MAR-1990

APPLICATION NUMBER: US 07/519,011

FILING DATE: 19-APR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Schwadron, Janet I.

REGISTRATION NUMBER: 33,778

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-708-1935

TELEFAX: 212-246-5959

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

; LENGTH: 492 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-07-783-705A-4

Query Match 6.9%; Score 191.5; DB 1; Length 492;

Best Local Similarity 23.9%; Pred. No. 1.5e-08;

Matches 131; Conservative 65; Mismatches 242; Indels 111; Gaps 26;

QY 9 VVSGHNALVSAAYLARBGMSVEVLEKDTVLGGAVSTVERPGYKVDGSSAHLMLRHSG 68

Db 6 VIGAGFGGALAIRLQAAGIPVLLLEQRDKPGRAVYVED-QGFTFDAGPT---VITDPS 61

QY 69 IIEEL-GLCAHGLR-YIDCDP-----WAPAPAPGTDGEGIVFHRDLQAT-CQSIER 117

Db 62 AIEELPALAGKQKKEYVELLPVTPFVRLCWE-----SGKVFNVDNDQTRLEAQIQ 111

QY 118 ACOTKADADAYRFAVWVSERSHYWK-----APSTPTGS--NLICAFGLATARG-NSEL 170

Db 112 QFNPRDVEGYRQFL-----DYSRAVFKEGYKLGTVFFLSFRDMLRAAPQLAKLQAWRSVY 167

QY 171 SRQFLAPGDALLDEYFDSALKAALAWFG-AOSGPPMSEPCTAPMVGFPAALAHVLPQGRA 229

Db 168 SK-----VASVIEDHLRQAFSHSLVGNPFATS-----SIYTLIHALERENG 212

QY 230 V-----GGGALSAALASMAVDGATVALGDGVTSTIRNSNH-WTVTTESGREVHARKVIA 284

Db 213 VWPFRGGTGAIVQGMIKLPQDLGGEVILNARVSHMETTGNKIEAVHLEDGRFRFLTAQAVAS 272

QY 285 GCHILNTL-DLLGNGGDFDRTTLQHWRRKIRVPGGIGAVLRATLSALPSYRGDATTREBS 343

Db 273 NADVHTYRDDLSQHPAAVQSNKLTQK-----RMSNSLFVLYFG----- 312

QY 344 GLQLVSDRAHLTAHGAALAGELPPPAVL-----GMS--FS-----GIDPTIA 386

Db 313 -----LNHHDLQAHHTVCFG---PRYRELIDEIFNHDGLAEDFSLYHLPACVTDSSIA 363

QY 387 PAGEHQVTLWSQVPLSGHRDOWASVAEAEADRVGEMEA-PAGFTDVLDRFIQTER 445

Db 364 PBGCGSYVYIAP-VPHLGATANLDWTVEGPKLDRIFAYLEQHFMGLRSQLTWTHMFTPF 422

QY 446 DIESELGMIGNVMHVMESLDQMLWRPLFELSGHRVPGADGLYLTCASTHPPGGVSGAS 505

Db 423 DFRDQLNAVHGSASFVEPVLTSAPWRP-----HNRDKTITNLVLVGAGTHGAGIPGVI 477

QY 506 GRSARIAL 514

Db 478 GSAKATAGL 486

## RESULT 10

US-09-934-903-16

; Sequence 16, Application US/09934903

; Patent No. 6660507

; GENERAL INFORMATION:

; APPLICANT: Koffas, Mattheos

; APPLICANT: Odom, J. Martin

; APPLICANT: Schenzle, Andreas J.

; APPLICANT: No. 6660507ton, Kelley C.

; APPLICANT: Tomb, Jean-Francois

; APPLICANT: Rouviere, Pierre

; APPLICANT: Picataggio, Stephen

; APPLICANT: Cheng, Qiong

; TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production

; FILE REFERENCE: CL1646 US NA

; CURRENT APPLICATION NUMBER: US/09/934,903

; CURRENT FILING DATE: 2001-08-22

; PRIOR APPLICATION NUMBER: 60/229,907

; PRIOR FILING DATE: September 1, 2001

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 16

LENGTH: 511  
TYPE: PRT  
ORGANISM: Methylobionas 16a  
FEATURE:  
OTHER INFORMATION: Amino acid sequences encoded by ORFs  
-09-934-903-16

Query Match 6.7%; Score 185; DB 4; Length 511;  
Best Local Similarity 21.1%; Pred. No. 5.8e-08;  
Matches 124; Conservative 87; Mismatches 226; Indels 150; Gaps 30;  
8 VVVGSGNALVSAAALAREGNSVEVLEKDTVLGGAVSTVERFPYKVDGSSAHLMIHRS 67  
8 IIVGAGGGLCAGMLLSORGFKVSIPDKHAEIGGRNPI--NNGFTDYGPFLM---X 63  
68 GIIEELGLGAGSLRYIDC-----DPWAFAPAG-----TDCPGIVFHRDLATCQSIERA 118  
64 GVLDME-----FELCRRSBDYLFELSLPMYRLLYDDRDIFVYSRENNRAELQEV 115  
119 CGTDADAYRRFVAVWESR-----HVMKAFSTPTGSGNLICAFGGLATARG 166  
116 F-DEGTGUYEQFME--QERKFNALYPCITRDYSSLSKF-----LSLDLIKALPWLAPKAS 168  
167 --NSELGRQFLAPDALLDYFQSEALKALAWFGAQS--PPMSFCTAPMVGFALMHEV 223  
169 VFNN-----LGOYFNQEKMLAFQCFQSKYLGNPWCECPALFTMLPY--LEHE 213  
224 LPPCRVGGSGSALSALASMAVDGATVALGDGVT--IRNSNHWVTWTSRGREYHARKV 282  
214 YGIYHVGGNLRNIAAMQAVIENGGEIHLNSELIESLIENGAAKGVKQHGABLRGDEV 273  
283 IAG---CHILTTDLGNGGDFRTTLDHWRKIRVGGIGAVLRATSALPSYRGD-ATT 338  
274 IINADFAHMTL-----VKPGV-----LKXYTPENLQK 302  
339 RE-STSGQLLVLS-DRALHTAHAALAGELPPRPVILGNKF----- 378  
303 REYSCSTFMYLGLDKTY-----DLPHHTIVFAKDYTNIRNFDNKTITDDFS 351  
379 -----SGIDPTIAPAGRHQVTLWSQWQVYELSGHSDWASVAEADRIVGEMEPAPGF 432  
352 FYVQNASASDLSAPACKSALYVLP-PPNDSG-LDQWQHCONVREQVLDLGA-RGLG 408  
433 TD-----SVLDRIOTPRDISELGMIGNVHVEMLSDQMLWRLPELSGHRVPGADGLY 489  
409 SDIRAHIECEKIITPQWTERDEHYVKGATFSLGKFSQKLYWRP-----HNRFEELANCY 463  
490 LTGASTHPGGVSG--ASGRSAARIALSDSRCK-----ASQWRRS 529  
464 LVGGGTHPGSLFTIYESARISAKL-ISQKRVRFKDIASHAWLKXA 509

## SULT 11

-09-934-903-18  
Sequence 18, Application US/09934903  
Patent No. 6660507

## GENERAL INFORMATION:

APPLICANT: Koffas, Mattheos  
APPLICANT: Odom, J. Martin  
APPLICANT: Schenzle, Andreas J.  
APPLICANT: No. 6660507ton, Kelley C.  
APPLICANT: Tomb, Jean-Francois  
APPLICANT: Rouviere, Pierre  
APPLICANT: Picataggio, Stephen  
APPLICANT: Cheng, Qiong

TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production

FILE REFERENCES: C11646 US NA

CURRENT APPLICATION NUMBER: US/09/934,903

CURRENT FILING DATE: 2001-08-22

PRIOR APPLICATION NUMBER: 60/229,907

PRIOR FILING DATE: September 1, 2001

NUMBER OF SEQ ID NOS: 24

SOFTWARE: Microsoft Office 97

## RESULT 12

US-09-443-184-55

Sequence 55, Application US/09443184A

Patent No. 6372431

## GENERAL INFORMATION:

APPLICANT: Cunningham, Mary Jane  
APPLICANT: Zweiger, Gary  
APPLICANT: Kaser, Matthew R.  
APPLICANT: Panzer, Scott  
APPLICANT: Seilhammer, Jeffrey J.  
APPLICANT: Yue, Henry  
APPLICANT: Baughn, Mariah  
APPLICANT: Azimzai, Valda  
APPLICANT: Lal, Preeti

TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS

FILE REFERENCE: PC-0007 US

CURRENT APPLICATION NUMBER: US/09/443,184A

CURRENT FILING DATE: 1999-11-19

NUMBER OF SEQ ID NOS: 138

SOFTWARE: PERL Program

Query Match 6.0%; Score 165.5; DB 4; Length 497;  
Best Local Similarity 23.3%; Pred. No. 3e-06;  
Matches 128; Conservative 75; Mismatches 210; Indels 137; Gaps 30;  
QY 8 VVVGSGNALVSAAALAREGNSVEVLEKDTVLGGAVSTVERFPYKVDGSSAHLMIHRS 67  
Db 10 IIVGAGGGLSALSILATAGFSVQLIEKNKVGKGNIMTK-DGFTFDLGPSI-LTPH- 66  
QY 68 GIIEEL--GLGAGSLRYID--CDP-WA-FAPPAPGTD-----GPGIV-- 103  
Db 67 -IPEALFTGAGKNMADYVQIKVEPHWRNFEFGSDVIDLCEDAETQRRELDKLPGGTAY 125  
QY 104 FHRDLAT--CQSIERACCTKDAAYRRFVAVWESRSHVMKAFSTPTGSGNLICAFGG 160  
Db 126 FQRLDYSKNLCTETETAGYFAKGLDGF-----W-----DLLFYGP 161  
QY 161 LATARGNSLSRQFLAFGDALLDEYFDSEALKALAWFGAQSQGPMSSEPGTAPMVGFAL 220  
Db 162 L-----RSLLSFDVFRSMDOQVRRFISDPKLVILNYFIKVG---SSPYDAP---AL 208  
QY 221 MHVLPGR-----AVGSGSALSALASMAVD--GATVALGDGVTISIRNSNHWVTTE 272  
Db 209 MLLFPYIQHYGLVYVGGMYGQAQAM-EKLAVELGVEIRLDAEVSEIQK-----Q 258  
QY 273 SGREHARKVIAGCHILTTDLGNGGDFRTTLDHWRKIRVGGIGAVLRATSALPSY 332  
Db 259 DGR-----ACAV-----KLANG--DVLPAIVVSNMEVIPAMEKLLRSASELKQM 302  
QY 333 RGDATTRESTSGQL--LVSDRAHLRTHAALAGELPPR-----PAVLGMS 377  
Db 303 Q--RPFPSCSGLVHLGVDRLYLPQLAHNFFYSDPREHFDVAFKSHRLSDDDTYLVA 359  
QY 378 FSGIDPTIAPAGRHQVTLW---SQWQVYELSGHSDWASVAEADRIVGEMEPAPGF 434  
Db 360 PKTDPQAQAGCEIILPHILOPDKLLTAEDYSALRE-----RVLVKLERM--GLTD 413  
QY 435 ---SVLDRFOTPRDISELGMIGNVHVEMLSDQMLWRLPELSGHRVPGADGLYLT 491  
Db 414 LRQHVITVEYWTPLDIQAKYYSNQSGSIYGVVAD-----REFKNLGFAPQRSSELSNLYFV 468  
QY 492 GASTHPGGGV 501  
Db 469 GGSVNPGGGM 478

SEQ ID NO 55

LENGTH: 610

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No. 6372431 1867333CD1

IS-09-443-184-55

Query Match 5.3%; Score 146; DB 4; Length 610;  
Best Local Similarity 23.6%; Pred. No. 0.00021;  
Matches 134; Conservative 73; Mismatches 239; Indels 122; Gaps 29;

5 LDVVVGGSHNALVSAAYLAREGMSVEVLEKOTVLGAVSTVERPFGYKVDGSSAHLMI 64  
68 LDVVVGGSGGLAAAILAKAGKRVLVLEQHTKAGGCCCTFGK-NGLEFFDTG-----I 120  
65 RHSGIHELGLGAGHGYIDCDPWAPPAPGTD-----GPG-----IVHRDLDTCSI 115  
121 HYIGRMEEGIGRFLDQITQEGDWAFLSSFPDIMVLESGNGRKEYPMYSEKAYIQGL 180  
116 ERACGTQDA--DAYRFVAVMSRSHVMKAFSTPTGNSNLIGAFGLATARGNSLSRQ 173  
181 KEKFPQSEAIIDKYIKLVKVVSSGAPHAIIKFLPLPVVQLLDRCCLLTRFSPFLQASTQ 240  
174 FLAPGDALDEYFDSALKAALAWFGAQS--PPMS-----EPGTAPMVGPA 218  
241 SLA---EVLQGLGASSELQAVLSYIIFTGVTPNHSAPSWHALLVNHVMKGGFYERGSS 297  
219 AL-MHVLPPGKAVGGSGSALAASLMAVDGATVAGDGVTSIRNSNHMTVTTSGRE- 276  
298 EIAFHTPIVIOAGGA-VLTKATVQSVLLDSAGKAG-----VSVKKGHEL 342  
277 --VEARVVIAGCHILTLDLLNGGSDRTTLDHWRKI-RVGGIGAVLRATLSALPSVR 333  
343 VNIYCPVWSNAGLFNTYHLLPG--NARCLPGVKQOGLGTVRPGAG-----MTSPFICLR 395  
334 GDATTR-----STSGIQLLVSRRAHLRTAHGAALAGELPPRPVAVLMSF-SGIDPTIAP 387  
396 G---TKEDLHPSTNYVYVYDMDQWERYVSMPEEAHEHPLLFAPPSAKDPTWED 452  
388 --AGRHQVTL-----WSQWQPYRLSGHR--DWA-----SVABEAD-----RIVCE 424  
453 RFGKSTMIIMLPTAYEWEFMOA-ELKGRGSDYETFXNSFVEASMSVVLKFLFQLECK 511  
425 MEAFAPGFTSVLDR--FIQTPR-----DIESELGMIGNNVHVMVMSLDQMLWRPLP 478  
512 VESWTAG---SPLTNQPYLAAPRGACYGADHDLCRLHPCVM-----ASLRQSPIP--- 559  
479 GHRVPGADGLYLTGASTHPGGVSGASG 506  
560 -----NLVLTGQDIFTGCLVQALQG 579

## RESULT 13

US-09-252-991A-19574

Sequence 19574, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 19574

LENGTH: 863

TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19574

Query Match 5.0%; Score 139; DB 4; Length 863;  
Best Local Similarity 22.4%; Pred. No. 0.0015;  
Matches 114; Conservative 51; Mismatches 187; Indels 156; Gaps 25;

87 PNAPAPPAPGT--DQPGIVFHRD-----LDATQOSTERACGTKDADAYRFRFVAVWSESRH 140  
36 PRFSPSTSAWPDGFGWFSNTPTSPISPSAGPIRSGSSTSTS-----AIIATWT----- 86  
141 VMKAFSTPTPGSNLIGAFG-GLATARGNSELGRQFLAPGDALLDEYFDSALKAAL---A 196  
87 -----TGT-----GANPCGLECSAGSSCRNRWHERPA-----ARRLGAGLRAP 127  
197 WFGAQSGPMSEPGTAPMVGFALMHVLPGRVAVGGSGALSALASMAVDGATVALGCG 256  
128 WLSAHVRRRRADPGVPFAVAF-----PGRHAGAAG--PGAVGRNRNSTGIRRLGTR 178  
257 VTSIRENSNHMTVTTSGREHV-----ARKVIAGCHILTLDLLNGGSDRT 302  
179 RTAVRPTLR---CRTSVRPVHLPLWLSGRADPPYRPGHQRGACPAFTGRP--GQGRGER 233  
303 TTLDMRRKIRVPGIGAVLRATLSALPSYRGDATTBSTSGIQLLVSRRAHLRTAHG--- 360  
234 -----RPVVRTPGAGPLRRPAGGVPRLVGA-----CAARRAVRPGPLP 275  
361 AALAGELPPRPVAVLGSFSGIDPTIA-----PAGRHQVTLMSQWQPYRLSGHRDWSV 413  
276 APTAARLPAR--LAARYAGRRRTVAASAGRRHPAGRHRL----- 313  
414 ABABADRIVE-----MEAFAPGFTSVLD---RFIQTPTDIESELGMIG 455  
314 -RAAAGGLAGTGGWLADLSGVDCRVATSPGREDPRLEDVHRGRVPPAHQDMAGPG 372  
456 GNNVHVMVMSLDQMLWRPL-----PELSGHRV-----PGADGLYLTGASTHPGGV 501  
373 GRPGVSAVAVARL--RPAHRAQRPDRAARMACGGRQPCRPQHPVLRQAGSSHPGTGR 430  
502 SGAS-----GRSAARIALSDSRGKAS 523  
431 GGAAAVPVPPGAAAGEHGLVADRFQGPS 458

## RESULT 14

US-09-252-991A-30706

Sequence 30706, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 30706

LENGTH: 538

TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30706

Query Match 4.8%; Score 134; DB 4; Length 538;  
Best Local Similarity 20.9%; Pred. No. 0.002;  
Matches 123; Conservative 73; Mismatches 187; Indels 206; Gaps 30;

5 LDVVVGGSHNALVSAAYLAREGMSVEVLEKDTVLGCAVSTVERPFGYKVDGSSAHLMI 64  
40 VDMLLVGAG---IMSATL-----AVLLKELDPNLKWE-----VVEL 72

```

65 RHSGIIEBGLGAHGLRVIDCDPAFAPAPGCTDGPVGFVHRDLDTACQSIERACGTGDA 124
73 QBSGAIE-----SSNPWNA-----GTGAGLIC-----ELNYPQSGADSIDIKKA 113
125 DAYRFRVAVMSERSFVVK--AFSTPTGSLNIGAFGLATARGNSLSRQFLAPGDALL 182
114 VCINTMFVSKQFWSHLVAKGTGSKTFINPVPHLSFVRGSEGIIVLKKRFS-----L 168
183 DRYFDSEAL-----KAALA-WFGAOSGPPMSEPGTAPMVGPAALMHVLPGRVAGGS--- 233
169 TGHAFETWYSEDKATLAEM-----PLMMPCKRPADEALAAAT-----RVGGGTDVN 215
234 GALSAAALASRMVD-GATVALGCGVTISRNSNHWTVT---TESG--REVHARKVIAGC 286
216 FGALTNQLOHLAEOPGACIRYNQKTHLRADNGRVTVDTRNGGDRETQARFV----- 271
287 HILTTLDLGNGFPORTTLDHRRKIRVGPVIGAVLRATLSALPSYRGDATTRESTGLQ 346
272 -----FLGAGG-----GALFLQLSGIPEGKG----- 293
347 LUVSRAHLRTAAGLAAGELPPRPVLCMSFGIDPTIAPAGRHQVTLASOWQ---PYR 403
294 -----FGGPP-----VSGWERCNDNPEIVK--CHQAKVYSQAEVGSPPM 330
404 LSGHRDWSVAEAEADRIVGRMEAFAPGFTDSVLDRTFQTPRDIIESELGIGNVM-HVE 462
331 SVPHLD-TRVDGKSLFLGPPVAGFS-----TKFLRHGSFLDLPLSVRPGNLPMLS 381
463 MELDOMQWRPL-----PELSGHRV-----PGADGL 488
382 VARDNNDLTRYLIGVQMSPEQRLEALRKFYFEARAEDWRLEVAGQRVQIIKDPKRGGI 441
489 YLTG---ASTHPG-----GGVSGASGRSAARIALSD---SRRGKASOW 525
442 LQFGTELVAHDGSIALLGASPGASVTIVSIMGIERCPEQARSPEW 490

```

## RESULT 15

3-09-036-987A-5

Sequence 5, Application US/09036987A

Patent No. 6143526

## GENERAL INFORMATION:

APPLICANT: Baltz, Richard H.  
 APPLICANT: Broughton, Mary C.  
 APPLICANT: Crawford, Kathryn P.  
 APPLICANT: Madduri, Krishnamurthy  
 APPLICANT: Merlo, Donald J.  
 APPLICANT: Treadway, Patti J.

APPLICANT: Turner, Jan R.

APPLICANT: Walldron, Clive

TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide

TITLE OF INVENTION: Production

NUMBER OF SEQUENCES: 39

CORRESPONDENCE ADDRESS:

ADDRESSER: Dow AgroSciences LLC Patent Department

STREET: 9330 Zionsville Road

CITY: Indianapolis

STATE: Indiana

COUNTRY: USA

ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/036,987A

FILING DATE: 09-MAR-1998

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Stuart, Donald R

REGISTRATION NUMBER: 28,479

REFERENCE/DOCKET NUMBER: 50,608

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317)337-4816

TELEFAX: (317)337-4847

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 4928 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-036-987A-5

Query Match 4.8%; Score 133; DB 3; Length 4928;

Best Local Similarity 23.4%; Pred. No. 0.062;

Matches 151; Conservative 61; Mismatches 196; Indels 238; Gaps 39;

```

QY 3 AFID--AVVVGSHNALVSAAYLAREGWSV-EVLEKDTVLCGAVSTV-----ERFPG--- 51
DB 531 AALDERAVVIGSDRAALLTLGLRAFADGCDAPFVSGSVGLGRGVFVPSGGGGQWPMGNGR 590
QY 52 --YKV-----DR-----GSSAHLMIR-----HSGIIE-ELQ-- 74
DB 591 GLYSVPVPADAFDEACAEALDAHLQELVRDVFVFGSQAWLLDRVTWVWQSGLFALQIGLL 650
QY 75 --LGAGHLYRIDCDPWAFAPAPGTDGPGIVFHRDLDTACQSIERACGTGADAYRRVA 132
DB 651 RLIGSWGVA-----PDVVLGHSVSEL--AAVHAAGVLSLSEAAARLVA 690
QY 133 VMSERSHVMKAFSTPTGSLNIGAFGLATARGNSLSRQFLAPGDALLDEYFDSALK 192
DB 691 ---GRAR-LAQAL---PSG---GAM--LAVATGEFQV-----DFLLD----- 720
QY 193 AALAWFGAOSGPPMSEPGTAPMVGPAALMHVLPGRVAVGSGSALSAAALASRMVDGATVA 252
DB 721 -----GVRDRIGIAAVNG--PESVLSGDRELLTEIADRLHDQGC-- 759
QY 253 LGDGVTSIR-----NSNHWTVTTES-----GREVHARKVIAGCHILTTDLGNGGPD 301
DB 760 -----TRWLRVSHAFESHPEMPELEPAQISRGREYHAPEL-----PIISTLGEHLDGGRV 810
QY 302 RTTLDHWRKIR-----VPGGIGAVLRATLSALPSYRGDATTRES--TSGQL 347
DB 811 MGTPEYVWVQVREPVRPAEGVQALVGGQVGTIVELGPDGALSTLVEECVABSGRVAGIPL 870
QY 348 LVSDRAHLRTAAGLAIA-----GELPPRPVAVLGMFSFGID-PTIAPAGRHQVTLWSQWQ 400
DB 871 MKRDRDEARTVL-AALAQIHTRGGEVDWRSFPFAGTGAKQVLDLPTVA-----FORQ 919
QY 401 PYELSG---HRDWSVAEAEADR-IWGMEAFAPGFTDSVLDRTFQTPRDIIESELGIMGG 456
DB 920 RYWLSTGRAGDVTAAGLAEDHPILGAVVALLADG-----EGVVLGR-----LTAG 966
QY 457 NVMHVEMSLDQMWLMRPLPELSGHRVPG---ADGLYLTGASTHPG----- 498
DB 967 S--H-----PWLSDHRVLGEIVVPGTAIVELVMHVGERLGCGRVEELEALEA 1010
QY 499 -----GGVS-----GASGRSAARIALSDSRGKA--SOWMRRSS 530
DB 1011 PLILPDHGAVOQVVLVGPPEGSARSAVALYSCPGAIEPWEKKHAT 1056

```

Search completed: February 29, 2004, 14:55:08

Job time : 19.1665 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

! protein - protein search, using sw model

in on: February 29, 2004, 14:26:38 ; Search time 62.8999 Seconds  
(without alignments)  
2389.754 Million cell updates/sec

file: US-09-941-947a-38

irect score: 2768

quence:

1 MSAPFLDAVVGVGSHALVSA.....ALSDSRGKASQMMRRSSRS 532

oring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : A.Geneseq\_29Jan04.\*

1: geneseqp1980s.\*

2: geneseqp1990s.\*

3: geneseqp2000s.\*

4: geneseqp2001s.\*

5: geneseqp2002s.\*

6: geneseqp2003as.\*

7: geneseqp2003bs.\*

8: geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2768	100.0	532	5	Aae22317 Rhodococc
2	2764	99.9	532	6	ABP71892 R. erythr
3	859.5	31.1	511	6	ABP71893 D. radiod
4	688	24.9	542	6	ABP71894 Synchpoc
5	624.5	22.6	581	4	AM93657 Human pol
6	624.5	22.6	581	5	ABP7177 Human oxi
7	403	14.6	504	7	ADC31532 Human nov
8	305.5	11.0	322	7	ADC3152 Human nov
9	296.5	10.7	471	4	AG90287 C glutami
10	270	9.8	501	6	AAO22615 Synchocy
11	225	8.1	582	2	AAW22499 Phaffia d
12	225	8.1	615	6	AAO22614 Tomato pl
13	221.5	8.0	548	4	AB85728 Enzyme in
14	219.5	7.9	548	4	AB76640 Corynebac
15	219.5	7.9	548	4	AB76641 Corynebac
16	219.5	7.9	548	4	AG90440 C glutami
17	217.5	7.9	494	2	AAW00871 Flavobact
18	214.5	7.7	494	2	AAW69532 Flavobact
19	200	7.2	411	2	AAW83358 Mycobacte
20	198.5	7.2	494	6	AM70123 Photorhab
21	195.5	7.1	544	5	AAO15518 Agromyces
22	192.5	7.0	530	6	AAE31690 Rhodococc
23	191.5	6.9	492	2	AAW07466 Polypepti
24	191.5	6.9	492	2	AAW82257 C. utilis
25	191.5	6.9	492	2	AAW87889 Protein e

26	191.5	6.9	492	2	AAW99099	Erwinia u
27	191.5	6.9	492	2	AAW26333	Erwinia u
28	189.5	6.8	492	5	AAE22314	Pantoea s
29	189.5	6.8	492	5	AAO16021	Pantoea s
30	189.5	6.8	492	6	ABP96688	Pantoea s
31	185	6.7	511	5	AAE22309	Methylo
32	185	6.7	511	5	ABG61588	High grow
33	185	6.7	511	5	AAU80332	Methylo
34	185	6.7	511	5	ADA14534	Methylo
35	182.5	6.6	502	6	ABU43877	Protein e
36	182	6.6	121	4	AAU86860	Novel hum
37	182	6.6	121	7	ADB60194	Connectiv
38	182	6.6	595	5	AB890884	Herbicida
39	181	6.5	587	6	AAO22616	Arabidops
40	173.5	6.3	205	4	AAW95800	Human rep
41	173.5	6.3	205	4	AB896331	Human tes
42	167.5	6.1	497	6	ABU16031	Protein e
43	167.5	6.1	526	2	AAW95697	Erythroba
44	165.5	6.0	497	5	AAE22310	Methylo
45	165.5	6.0	497	5	ABG61589	High grow

#### ALIGNMENTS

RESULT 1  
AAE22317  
ID AAE22317 standard; protein; 532 AA.

XX

AC AAE22317;

XX

DT 25-JUL-2002 (first entry)

XX

DE Rhodococcus erythropolis AN12 beta-carotene ketolase (CrtO) enzyme.

XX

XX Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet; anti-oxidant; steroid; flavour; fragrance; electro-optic application; aquaculture; enzyme; beta-carotene ketolase; CrtO.

XX

OS Rhodococcus erythropolis.

XX

FH Key Location/Qualifiers

FT Misc-difference 1

FT /note= "Encoded by GTG"

XX

PN WO200218617-A2.

XX

PD 07-MAR-2002.

XX

PF 04-SEP-2001; 2001WO-US027420.

XX

PR 01-SEP-2000; 2000US-0229858P.

XX

PR 01-SEP-2000; 2000US-0229907P.

XX

PA (DUPO ) DU PONT DE NEMOURS & CO B I.

XX

PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX

PI Odom JM, Picataggio SK, Rouviere PE;

XX

DR WPI; 2002-351711/38.

XX

DR N-PSDB; AAD35515.

XX

PT Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by using microorganisms having a nucleic acid molecule encoding enzymes in the carotenoid biosynthetic pathway and which metabolize single carbon substrates.

XX

PS Claim 27; Page 148-150; 156pp; English.

XX

CC The invention relates to a method for producing carotenoid compounds. The method comprises a transformed metabolising host cell, comprising suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule encoding an enzyme in the carotenoid biosynthetic pathway, under the

C control of regulatory sequences, and contracting the host cell with carbon  
 C substrate to produce a carotenoid compound. The method is useful for  
 C producing carotenoid compounds such as anthraxanthin and astraxanthin, by  
 C using microorganism having a nucleic acid molecule encoding enzymes in  
 C the carotenoid biosynthetic pathway and which metabolise single carbon  
 C substrates. The carotenoids have potent anti-oxidant properties useful in  
 C diet, and aquaculture elements. The carotenoids are also useful as  
 C intermediates in the synthesis of steroids flavours and fragrances and  
 C compounds for potential electro-optic applications. The present sequence  
 C is Rhodococcus erythropolis AN12 beta-carotene ketolase (CrtO) enzyme  
 C used in the invention

X Sequence 532 AA;

Query Match 100.0%; Score 2769; DB 5; Length 532;

Best Local Similarity 100.0%; Pred. No. 5.7e-223;

Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 MSAPLDVVVSGHNALVSAAYLAREGWSVEVLKDTVLGGAVSTVERPGYKVDGSSA 60  
 b 1 MSAPLDVVVSGHNALVSAAYLAREGWSVEVLKDTVLGGAVSTVERPGYKVDGSSA 60  
 Y 61 HLMIRHSGIIEELGAGHLYIDCDPWAFAPAGTDPGIVFHRDLATCQSIERACG 120  
 b 61 HLMIRHSGIIEELGAGHLYIDCDPWAFAPAGTDPGIVFHRDLATCQSIERACG 120  
 Y 121 TKDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGLATARGNSLSRQFLAPGDA 180  
 b 121 TKDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGLATARGNSLSRQFLAPGDA 180  
 Y 181 LLDYFDSALKAALAWFGAQSPPMSEPGTAPWVGFPAALMHVLPGRVAGSGALSAAL 240  
 b 181 LLDYFDSALKAALAWFGAQSPPMSEPGTAPWVGFPAALMHVLPGRVAGSGALSAAL 240  
 Y 241 ASRMADGATVALGCGVTSIRNSNHNVTTESGREVHARKVIAGCHILTLDLGNGGF 300  
 b 241 ASRMADGATVALGCGVTSIRNSNHNVTTESGREVHARKVIAGCHILTLDLGNGGF 300  
 Y 301 DRTTLHWRKIRVPGIGAVLRATLSALPSYRGDATTRESTSGQLLVSDRAHLRTAHG 360  
 b 301 DRTTLHWRKIRVPGIGAVLRATLSALPSYRGDATTRESTSGQLLVSDRAHLRTAHG 360  
 Y 361 AALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLMSQWOPYRLSGHRDWSVAEAEADR 420  
 b 361 AALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLMSQWOPYRLSGHRDWSVAEAEADR 420  
 Y 421 IVGEMEAPAGTDSVLDRTIOTPRDIESTBELGMIGNMVHVSMDQMMLWRPLPGLSGH 480  
 b 421 IVGEMEAPAGTDSVLDRTIOTPRDIESTBELGMIGNMVHVSMDQMMLWRPLPGLSGH 480  
 Y 481 RVPAGDGLYLTGASTHFGGVSAGSRAARIALSDSRKASQWNRSSRS 532  
 b 481 RVPAGDGLYLTGASTHFGGVSAGSRAARIALSDSRKASQWNRSSRS 532

RESULT 2

ABP71892

ABP71892 standard; protein; 532 AA.

AC ABP71892;

CT 10-MAY-2003 (first entry)

DE R. erythropolis carotenoid ketolase crtO.

GW Carotenoid ketolase; enzyme; cyclic ketocarotenoid biosynthesis; crtO.

JS Rhodococcus erythropolis.

NN WO2003012056-A2.

SD 13-FEB-2003.

XX

PF 01-AUG-2002; 2002WO-US024317.  
 XX  
 PR 02-AUG-2001; 2001US-0309653P.  
 XX  
 PA (DUPO) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Cheng Q, Tao L;  
 XX  
 DR WPI; 2003-300493/29.  
 XX  
 DR N-PSDB; ABZ75460.  
 XX  
 PT New isolated nucleic acid encoding carotenoid ketolase enzyme, useful for  
 PT producing cyclic ketocarotenoid compounds such as adonirubin, echinenone,  
 PT and as probes or primers to identify nucleic acids encoding the enzyme.  
 XX  
 PS Claim 4; Page 65-67; 90pp; English.  
 XX  
 CC The invention relates to a novel isolated nucleic acid encoding a  
 CC carotenoid ketolase enzyme which: (i) encodes an amino acid sequence  
 CC containing all six conserved motifs of CrtO enzymes of Rhodococcus  
 CC erythropolis AN12 strain, Deinococcus radiodurans R1 strain, and  
 CC Synechocystis sp. PCC6803 strain; (ii) encodes a sequence of 532 amino  
 CC acids, given in specification; or (iii) hybridises to (i) or (ii). A  
 CC nucleic acid of the invention is useful for obtaining a nucleic acid  
 CC encoding a carotenoid ketolase enzyme. A nucleic acid of the invention is  
 CC also useful for producing cyclic ketocarotenoid compounds. A chimeric  
 CC gene is useful for regulating cyclic ketocarotenoid biosynthesis in an  
 CC organism by introducing the gene into a host cell and growing the host  
 CC cell under conditions whereby the carotenoid ketolase gene is expressed  
 CC and the cyclic ketocarotenoid biosynthesis is regulated. The regulation  
 CC may be upregulation of cyclic ketocarotenoid biosynthesis, where the  
 CC carotenoid ketolase gene is overexpressed on a multicopy plasmid or is  
 CC operably linked to a inducible or regulated promoter. Optionally the  
 CC cyclic ketocarotenoid biosynthesis may be down regulated, where the  
 CC carotenoid ketolase gene is expressed in antisense orientation or is  
 CC disrupted by insertion of foreign DNA into the coding region. The present  
 CC sequence represents the R. erythropolis strain AN12 crtO polypeptide  
 XX

SQ Sequence 532 AA;

Query Match 99.9%; Score 2764; DB 6; Length 532;

Best Local Similarity 99.8%; Pred. No. 1.2e-222;

Matches 531; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSAPLDVVVSGHNALVSAAYLAREGWSVEVLKDTVLGGAVSTVERPGYKVDGSSA 60  
 DB 1 VSAPLDVVVSGHNALVSAAYLAREGWSVEVLKDTVLGGAVSTVERPGYKVDGSSA 60  
 QY 61 HLMIRHSGIIEELGAGHLYIDCDPWAFAPAGTDPGIVFHRDLATCQSIERACG 120  
 DB 61 HLMIRHSGIIEELGAGHLYIDCDPWAFAPAGTDPGIVFHRDLATCQSIERACG 120  
 QY 121 TKDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGLATARGNSLSRQFLAPGDA 180  
 DB 121 TKDADAYRRFVAVWSESRHVMKAFSTPTGSLNIGAFGLATARGNSLSRQFLAPGDA 180  
 QY 181 LLDYFDSALKAALAWFGAQSPPMSEPGTAPWVGFPAALMHVLPGRVAGSGALSAAL 240  
 DB 181 LLDYFDSALKAALAWFGAQSPPMSEPGTAPWVGFPAALMHVLPGRVAGSGALSAAL 240  
 QY 241 ASRMADGATVALGCGVTSIRNSNHNVTTESGREVHARKVIAGCHILTLDLGNGGF 300  
 DB 241 ASRMADGATVALGCGVTSIRNSNHNVTTESGREVHARKVIAGCHILTLDLGNGGF 300  
 QY 301 DRTTLHWRKIRVPGIGAVLRATLSALPSYRGDATTRESTSGQLLVSDRAHLRTAHG 360  
 DB 301 DRTTLHWRKIRVPGIGAVLRATLSALPSYRGDATTRESTSGQLLVSDRAHLRTAHG 360  
 QY 361 AALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLMSQWOPYRLSGHRDWSVAEAEADR 420  
 DB 361 AALAGELPPRAVLGMSFGIDPTIAPAGRHQVTLMSQWOPYRLSGHRDWSVAEAEADR 420  
 QY 421 IVGEMEAPAGTDSVLDRTIOTPRDIESTBELGMIGNMVHVSMDQMMLWRPLPGLSGH 480



|||||  
 421 IVGMBAPAGCTDSVLDRTIQTPTDIESELGIGNVHEVMSLQMMWRPLPLSGH 480  
 481 RVPAGDGLYLTGASTHPGGVSGASGRSAARIALSDSRKASQWNRSSRS 532  
 482 RVPAGDGLYLTGASTHPGGVSGASGRSAARIALSDSRKASQWNRSSRS 533  
 483 RVPAGDGLYLTGASTHPGGVSGASGRSAARIALSDSRKASQWNRSSRS 534

## SULT 3

P71893

ABP71893 standard; protein; 511 AA.

ABP71893;

10-MAY-2003 (first entry)

D. radiodurans carotenoid ketolase crtO.

Carotenoid ketolase; enzyme; cyclic ketocarotenoid biosynthesis; crtO.

Deinococcus radiodurans.

WO2003012056-A2.

13-FEB-2003.

01-AUG-2002; 2002WO-US024317.

02-AUG-2001; 2001US-0309653P.

(DUPO ) DU PONT DE NEMOURS &amp; CO E I.

Cheng Q, Tao L;

WPI; 2003-300493/29.

N-PSDB; ABZ75461.

New isolated nucleic acid encoding carotenoid ketolase enzyme, useful for producing cyclic ketocarotenoid compounds such as adonirubin, echinenone, and as probes or primers to identify nucleic acids encoding the enzyme.

Claim 15; Page 69-71; 90pp; English.

The invention relates to a novel isolated nucleic acid encoding a carotenoid ketolase enzyme which: (i) encodes an amino acid sequence containing all six conserved motifs of CrtO enzymes of Rhodococcus erythropolis AN12 strain, Deinococcus radiodurans R1 strain, and Synecocystis sp. PCC6803 strain; (ii) encodes a sequence of 532 amino acids, given in specification; or (iii) hybridizes to (i) or (ii). A nucleic acid of the invention is useful for obtaining a nucleic acid encoding a carotenoid ketolase enzyme. A nucleic acid of the invention is also useful for producing cyclic ketocarotenoid compounds. A chimeric gene is useful for regulating cyclic ketocarotenoid biosynthesis in an organism by introducing the gene into a host cell and growing the host cell under conditions whereby the carotenoid ketolase gene is expressed and the cyclic ketocarotenoid biosynthesis is regulated. The regulation may be upregulation of cyclic ketocarotenoid biosynthesis, where the carotenoid ketolase gene is overexpressed on a multicopy plasmid or is operably linked to a inducible or regulated promoter. Optionally the cyclic ketocarotenoid biosynthesis may be down regulated, where the carotenoid ketolase gene is expressed in antisense orientation or is disrupted by insertion of foreign DNA into the coding region. The present sequence represents the D. radiodurans strain R1 crtO polypeptide

Sequence 511 AA;

Query Match

Best Local Similarity 31.1%; Score 859.5; DB 6; Length 511;

Matches 207; Conservative 80; Mismatches 208; Indels 27; Gaps 11;

6 DAVVSGSHNLVSAAYLARGWGVLEKDTVLGGAVSVTERFPGVKYDRSSAHLMLR 65

5 DLIVMGAGNALVTAAAYARAGLVGVFERRHLVGGAVSTEEVVPVGRFDYGGSAHLIR 64

QY 66 HSGIIEELGAGHLIRYIDCDWAFAPPAAGTGGPIVF-HRDLDTAQSI-BRACGTKD 123  
 DB 65 MTPIVRELETRHGLHYLEVDPMFHA-----SGETPWFJHRDAGRITRELDKFPG--Q 117  
 QY 124 ADAYERFVAVKMSRSRHRVMAFSTPTTGSNLIAGFGGLATARG-----NSLSRQFLAPG 178  
 DB 118 GDVGRFLDDWTFFARAVADLNSAPGFLDL----GKVMVRSQGQKDWNEQLPR-ILRPY 172  
 QY 179 DALLDEYFDSEALKAAALWFAQSGPPMSEPGTAPMVGFPAALAHVLPPGRAVGGSGALSA 238  
 DB 173 GDVAREYFSEERVAEPLTWAAQSGPPSPDLGAPFLIMHPLHYHEGVARPKGSGGLTK 232  
 QY 239 ALASRMAVDGATVALGDGVTISIR-RNSNHWITVTTESGREVHARKVIAGCHILTTDLIGN 297  
 DB 233 ALRRATEABEGGEVETDAPVKEILLVKCKQAQIRLESGETVTARAVSVGHILTTANAL-- 290  
 QY 298 GGFORTWLDHWRKIRVPGPIGAVIRLATSALSYRGDATTRESTGQLIIVSDRAHLRT 357  
 DB 291 ---PAEYVPSAARNVRVNGFGMILRLALSERVKYR-HHTEPDSRIGLGLLIKNEQIMQ 346  
 QY 358 AFGAALAGELPPRPRAVLGMSFGIDPTIAPAGRHQVTLMSQWQPYRLSGHRDWSVAEAE 417  
 DB 347 GYGYLAGQPTTDEPLVAMSFSAVDDSLAPPNGDVLWMAQYYPFELA-TGSMETRTABA 405  
 QY 418 ADRIVGEXEAPAGFTSDSVLDRTIQTPTDIESELGIGNVHEVMSLQMMWRPLPEL 477  
 DB 406 RENTLRAFEHYAPGTRDTIVGELVQTPQWLETNLGLHGRGNVMELEMSFDQMFSPRWLKA 465  
 QY 478 SCHRVPGADGLYLTGASTHPGGVSGASGRSAARIALSDSR 519  
 DB 466 SQYRWPGVQGLYLTGASTHPGGVSGASGRSAARIVKDLTR 507

## RESULT 4

ABP71894

ID ABP71894 standard; protein; 542 AA.

XX ABP71894;

XX 10-MAY-2003 (first entry)

XX Synecocystis carotenoid ketolase crtO.

XX Carotenoid ketolase; enzyme; cyclic ketocarotenoid biosynthesis; crtO.

XX Synecocystis sp.

XX WO2003012056-A2.

XX 13-FEB-2003.

XX 01-AUG-2002; 2002WO-US024317.

XX 02-AUG-2001; 2001US-0309653P.

XX (DUPO ) DU PONT DE NEMOURS &amp; CO E I.

XX Cheng Q, Tao L;

XX WPI; 2003-300493/29.

XX N-PSDB; ABZ75462.

New isolated nucleic acid encoding carotenoid ketolase enzyme, useful for producing cyclic ketocarotenoid compounds such as adonirubin, echinenone, and as probes or primers to identify nucleic acids encoding the enzyme.

XX Disclosure; Page 72-74; 90pp; English.

XX The invention relates to a novel isolated nucleic acid encoding a carotenoid ketolase enzyme which: (i) encodes an amino acid sequence containing all six conserved motifs of CrtO enzymes of Rhodococcus erythropolis AN12 strain, Deinococcus radiodurans R1 strain, and

*Synechocystis* sp. PCC6803 strain; (iii) encodes a sequence of 532 amino acids. Given in specification; or (iii) hybridises to (i) or (ii). A nucleic acid of the invention is useful for obtaining a nucleic acid encoding a carotenoid ketolase enzyme. A nucleic acid of the invention is also useful for producing cyclic ketocarotenoid compounds. A chimeric gene is useful for regulating cyclic ketocarotenoid biosynthesis in an organism by introducing the gene into a host cell and growing the host cell under conditions whereby the carotenoid ketolase gene is expressed and the cyclic ketocarotenoid biosynthesis is regulated. The regulation may be upregulation of cyclic ketocarotenoid biosynthesis, where the carotenoid ketolase gene is overexpressed on a multicopy plasmid or is operably linked to a inducible or regulated promoter. Optionally the cyclic ketocarotenoid biosynthesis may be down regulated, where the carotenoid ketolase gene is expressed in antisense orientation or is disrupted by insertion of foreign DNA into the coding region. The present sequence represents the *Synechocystis* sp. strain PCC6803 crto polypeptide sequence 542 AA;

Q Sequence 542 AA;  
Query Match 24.9%; Score 688; DB 6; Length 542;  
Best local Similarity 31.9%; Pred. No. 9.1e-49;  
Matches 175; Conservative 97; Mismatches 229; Indels 48; Gaps 16;

[illegible]

RESULT 5  
AM93657

RESULT 5  
VAM93657

D AAM93657 standard; protein; 581 AA.

100

VC  
AAM

TX	06-NOV-2001	(first entry)
----	-------------	---------------

Human polypeptide, SEQ ID NO: 3528.

XX	Human, full length cDNA; cDNA synthesis; oligo-capping.
XX	
XX	Homo sapiens.
XX	
XX	EP1130094-A2.
XX	
XX	05-SEP-2001.
XX	
XX	07-JUL-2000; 2000EP-00114089.
XX	
XX	08-JUL-1999; 99JP-00194486.
XX	11-JAN-2000; 2000JP-00118774.
XX	02-MAY-2000; 2000JP-00183765.
XX	
XX	(HELI-) HELIX RES INST.
XX	
XX	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX	
XX	WPI: 2001-524255/58.
XX	N-PSDB; AAK94592.
XX	
XX	830 Primers useful for synthesizing full length cDNA clones and their use
XX	in genetic manipulation.
XX	
XX	Claim 8; SEQ ID NO 3528; 1380pp + Sequence Listing; English.
XX	
XX	The invention relates to primers for synthesising full length cDNA
XX	clones. 830 cDNA molecules encoding a human protein have been isolated
XX	and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
XX	been determined. Primers for synthesising the full length cDNA are useful
XX	for clarifying the function of the protein encoded by the cDNA. The full
XX	length clones were obtained by construction of full length enriched cDNA
XX	libraries that were synthesised by the oligo-capping method. The primers
XX	enable the production of the full length cDNA easily without any special
XX	methods. The present sequence is a polypeptide encoded by a full length
XX	human cDNA of the invention. Note: The sequence data for this patent did
XX	not form part of the printed specification, but was obtained in CD-ROM
XX	format directly from EPO
XX	
XX	Sequence 581 AA;
XX	
XX	Sequence 581 AA;

[illegible]

[illegible]

## RESULT 7

ADC31532  
ID ADC31532 standard; protein; 504 AA.

AC ADC31532;  
XX  
DT 18-DEC-2003 (first entry)

Human novel polypeptide sequence, SEQ ID NO:1614.

KW biodiversity assessment; parkinson's disease; alz  
KW neurodegenerative diseases; anaemia; platelet dis

ulcers; osteoporosis; autoimmune disease; cancer;  
 molecular weight marker; food supplement; antiparkinsonian; neurotropic;  
 neuroprotective; anti-anemic; anticoagulant; thrombolytic; vulnerary;  
 antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
 gene therapy; chromosome 10.

FF 24-SRP-2002; 2002WO-US030474.

PR 24-SEP-2001; 2001US-0324631P.  
XX  
PA (HYSE-) HYSEO INC.

XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac RT;  
XX WPI: 2003-371981/35.  
XX N-PSDB: ADC30561.  
XX New polynucleotide and polypeptide useful for diagnosing, preventing or  
XX treating conditions such as neurodegenerative diseases, anemias, platelet  
XX disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
XX cancer.  
XX Claim 20; SEQ ID NO 1614; 1185pp; English.  
XX The invention relates to 971 novel human cDNA sequences (ADC29919-  
XX ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
XX invention also relates to nucleic acid sequences over 99% identical with  
XX the novel human cDNAs. The invention additionally encompasses expression  
XX vectors and host cells comprising a nucleic acid of the invention; the  
XX recombinant production of a polypeptide of the invention; an antibody  
XX against a polypeptide of the invention; a method of detecting  
XX polynucleotides or polypeptides of the invention; and methods of  
XX identifying a compound which binds to a polypeptide of the invention. The  
XX invention further discloses methods of preventing, treating or  
XX ameliorating a medical condition; kits comprising polynucleotide probes  
XX and/or monoclonal antibodies for carrying out the methods of the  
XX invention; methods for the identification of compounds that modulate the  
XX expression or activity of the polynucleotide and/or polypeptide; and 767  
XX contig sequences corresponding to the cDNA sequences of the invention  
XX (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628  
XX -ADC33394). The nucleic acids and polypeptides of the invention are  
XX useful in diagnostics, drug screening, forensics, gene mapping, in the  
XX identification of mutations responsible for genetic disorders or other  
XX traits, for assessing biodiversity, and in producing many other types of  
XX data and products dependent on DNA and amino acid sequences. They are  
XX also used for treating diseases such as Parkinson's disease, Alzheimer's  
XX disease and other neurodegenerative diseases, anaemia, platelet  
XX disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
XX cancer. The nucleic acids may also be used as hybridisation probes or  
XX primers, and in the recombinant production of a protein. The polypeptides  
XX are also useful in generating antibodies, as molecular weight markers,  
XX and as food supplements. The present sequence represents a specifically  
XX claimed human polypeptide sequence of the invention. Note: The sequence  
XX data for this patent did not form part of the printed specification, but  
XX was obtained in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 504 AA;  
XX Query Match 14.6%; Score 403; DB 7; Length 504;  
XX Best Local Similarity 29.2%; Pred. No. 6.1e-25;  
XX Matches 140; Conservative 71; Mismatches 200; Indels 68; Gaps 17;  
2Y 6 DAVVVGSHNLAVALAREGNSVEVLEKDTVLGGAVSTVERPPGVKVDGSSAHLMR 65  
36 DAVVIGAGHGLVAALQRLGVNTAVFERHHVTHVGGAAVTEIIPGKFRASLYLSLLR 95  
2Y 66 HSGIIEELGAGHGLRYIDCPMAFAPAPGCTDGP-----IVFHRDLDTCCSIERAC 119  
2Y 96 PQ-IYTDLELXKGLRLHLPYFSTPL--EEGAGSKVPRCLLIGTDAENKQIAQ-? 151  
2Y 120 GTKDADAVRRVAVKNSRSHVMKAFSTPTPG-----SNLIGAFGLATARG----- 166  
2Y 152 SQDAQVPPKVEEFNHLALAIPELLDAAFPVDMAAFQHSLLQMRSLSTLKLKAGRI 211  
2Y 167 -NSELRF--LAPGDALLVEYDSIALKAALAW----FGAQSPPMSEGTAPMVGFAA 219  
2Y 212 LQAGLPRIEVLTPITKVLQWFESEPKATLATDAVIGNWSF--HTGS---GYVL 265  
2Y 220 LMHVLP-----GRAVGGSGALSAALASRMVADGATVALGCVGTSIRNSNHW--V 269  
2Y 266 LHHVVGLEGMQGWGANGYVQGGMGALSDAIASSATHTGASIFTEKTVAKVQVNSEGCVQ 325

QY 270 TTSGRVRHARKVIAGCH-----ILTTDLLNGSGFDRITLDWRKRVGPGICAVLR 323  
Db 326 VLEDGTEVRSMVLSNTSPQITFKLTPOEMLPEEFLEISQLDTSFV-----TKIN 378  
QY 324 LATSALPSY-----RGDATTRETSGLQLLVSDRAHLRTAAGALAGELPPRPVILGM 376  
Db 379 VAVDRLPFLAANPAGQPLPHQCS-IHLNCEDTLLHQAPEFADMGDLPSPHRPVIELC 437  
QY 377 SFGSIDTIPAGRHQVTLMSQVQVRLSGHRDW-ASVABEADRVIGENAEAPAGFTD 434  
Db 438 IPSSLDTLAPPGCHVVSLEFQYTPYTLAGKAWDEQERDAYADRVDFDCIEVYAPGKE 496  
RESULT 8  
ADC33152  
ID ADC33152 standard; protein; 322 AA.  
XX AC  
XX ADC33152;  
XX 18-DEC-2003 (first entry)  
XX Human novel contig-encoded polypeptide sequence, SEQ ID NO:3234.  
XX Human; diagnostic; drug screening; forensics; gene mapping;  
XX biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
XX neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
XX ulcers; osteoporosis; autoimmune disease; cancer;  
XX molecular weight marker; food supplement; antiparkinsonian; nootropic;  
XX neuroprotective; antianemic; anticoagulant; thrombolytic; vulnery;  
XX antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;  
XX gene therapy; chromosome 10.  
XX Homo sapiens.  
XX WO2003029271-A2.  
XX 10-APR-2003.  
XX 24-SEP-2002; 2002WO-US030474.  
XX 24-SEP-2001; 2001US-0324631P.  
XX (HYSE-) HYSEQ INC.  
XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
PI Haley-Vicente D, Drmanac RT;  
XX WPI: 2003-371981/35.  
XX N-PSDB: ADC32385.  
XX New polynucleotide and polypeptide useful for diagnosing, preventing or  
XX treating conditions such as neurodegenerative diseases, anemias, platelet  
XX disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
XX cancer.  
XX Example 2; SEQ ID NO 3234; 1185pp; English.  
XX The invention relates to 971 novel human cDNA sequences (ADC29919-  
XX ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The  
XX invention also relates to nucleic acid sequences over 99% identical with  
XX the novel human cDNAs. The invention additionally encompasses expression  
XX vectors and host cells comprising a nucleic acid of the invention; the  
XX recombinant production of a polypeptide of the invention; an antibody  
XX against a polypeptide of the invention; a method of detecting  
XX polynucleotides or polypeptides of the invention; and methods of  
XX identifying a compound which binds to a polypeptide of the invention. The  
XX invention further discloses methods of preventing, treating or  
XX ameliorating a medical condition; kits comprising polynucleotide probes  
XX and/or monoclonal antibodies for carrying out the methods of the  
XX invention; methods for the identification of compounds that modulate the  
XX expression or activity of the polynucleotide and/or polypeptide; and 767

contig sequences corresponding to the cDNA sequences of the invention (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628-ADC33394). The nucleic acids and polypeptides of the invention are useful in diagnostics, drug screening, forensics, gene mapping, in the identification of mutations responsible for genetic disorders or other traits, for assessing biodiversity, and in producing many other types of data and products dependent on DNA and amino acid sequences. They are also used for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, anaemia, platelet disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or cancer. The nucleic acids may also be used as hybridisation probes or primers, and in the recombinant production of a protein. The polypeptides are also useful in generating antibodies, as molecular weight markers, and as food supplements. The present sequence represents a human contig-encoded polypeptide sequence used in an example of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 322 AA;

Query Match 11.0%; Score 305.5; DB 7; Length 322;  
Best Local Similarity 30.5%; Pred. No. 4.9e-17;  
Matches 101; Conservative 34; Mismatches 117; Indels 79; Gaps 11;  
227 GRAVGGGSAALASRAVMDGATVALGDGVTISRRNSHWT--VTTESGREYHARKVIA 284  
26 GYVQGGGALSADASSATTHGASIFTEKTVAKVQVNSEGCVGVLEDGTEVRSKWL 85  
285 GCHILTLDLLNGGFFRTLDHRRKIRVGGIGAVLRATSALP-SYRGDATTRESIS 343  
86 NTSFQIT-----FLKLTQEW-----LPEFLERISOLDTRS 117  
344 GILQLVSDRAHLRTARGAALA-----GELPPPAVLGMSFGIDPTIAP-----AGR 390  
118 PVTKINXEAR---HIALSLPTEHSEKPGWQHLSHLLHCPDLQVPSCLYRSGR 173  
391 HQVTLMSQWQYRLSG-----HROMASVAAEAEDRIVGEM 425  
174 RQAQOPS-WRPPLPGASRCPIITNAPSTXVTKTSPSIFRLKPMWMACLPTVDCI---- 228  
426 EAFAPGFTDVLDFIOTFRDIESELGMIGNVHVMESLDQMLRPLPELSCHRVPGA 485  
229 EYVAPGFKDSVVGDLITPPDLERIFGLPGNIFHCAMSLDQLYFARVPLHSGIRCP-L 297  
486 DELYLTGASTHPGGVSGASGSAARIALSD 516  
288 QGLYLCSGSAHPGGVGMGAGRNAHVAFRD 318

SUIT 9

AG90287  
> AAG90287 standard; protein; 471 AA.

AAG90287;

26-SEP-2001 (first entry)

C glutamicum protein fragment SEQ ID NO: 4041.

Corynebacterium; amino acid synthesis; vitamin; saccharide;  
organic acid synthesis.

Corynebacterium glutamicum.

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-00127668.

16-DEC-1999; 99JP-00377484.

07-APR-2000; 2000JP-00159162.

PR 03-AUG-2000; 2000JP-00280988.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX WPI: 2001-376931/40.  
XX N-PSDB; AAH65506.

Novel polynucleotides derived from Corynebacterium bacteria, for identifying mutation point of a gene, measuring expression of a gene, analyzing expression profile or pattern of a gene and identifying homologous gene.  
Claim 17; SEQ ID NO 4041; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium, measuring expression amount and analyzing the expression profile or expression pattern of a gene derived from Corynebacterium, and identifying a homologue of a gene derived from Corynebacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office

XX Sequence 471 AA;

Query Match 10.7%; Score 296.5; DB 4; Length 471;  
Best Local Similarity 25.9%; Pred. No. 4.6e-16;  
Matches 141; Conservative 61; Mismatches 224; Indels 119; Gaps 23;  
QY 7 AVVVGSGHNAVSAYLAREGVSVELEKQVTLGASVTVERP-PGYKVDGSSAHLMI 65  
DB 13 AVVVGSGNGHUTTAVALAKAGQVDVYEAAPTGAARSESVLGEITISLGAHGFEGV 72  
QY 66 HSGIIEELGAGHGLRYIDCFNAPAPPA---PQTDGFGIVFHRDLDTQCSIERACGTX 122  
DB 73 ASPAPHYLGLEDHGLE-----WAYSPPMAHPLDYGRAGLETSLPETAKKL----- 119  
QY 123 DADAYRRFVAVWSERSRHVKAFSTPTPGNSLIG-----AFGLATARGNSE 169  
DB 120 GPDA-RRWKNLHQGLTKNIDKHL-----ANLGPVLKWAHPAHPMAKPGPFA----- 165  
QY 170 LSRQFLAPGDALDDEYFDSEALKAAALAWFGAQSGPPMSEPCAPMVGFPAALMHVLPFGR- 228  
DB 166 -----LLPAKLASAAFEETEARSIFICSAMHSVTPPHKPMWASL---GLLFGALGWSRG 217  
QY 229 ---AVGGGALSALASRAVMDGATVALGDGVTISRRNSHWTVTTESGREYHARKVIAG 285  
DB 218 WFAVVGSGRIVDALWNVINHHGGTIHCDSDISL-----SQFRDQDA--IILN 264  
QY 286 CHILTLDLLIG---NGGPDRTTLDHRRKIRVGGIGAVLRATSALP---SYRGDATT 339  
DB 265 QTPSQVLKLTGDLNAGLPQ-RMSTNKH---GFSSYKVDYLLDGPWPNSPQVQATT 319  
QY 340 ESTGLQLLVSDRAHLRTAHGAALAGBLPPPAVLGMSFGIDPTIAPAGRHQVTLMSQW 399  
DB 320 H-----VGGSSSEIAPAEAAVAAGRMPEFPFIILCQQQVADPSRAREGSHVY--WA-- 368  
QY 400 QFYRLSGHRDWSVAEAEADR-----IVGMEAFAPGTDVSLDRFIQTREDIES-BELGM 453  
DB 369 -----YAVPRGFDVKRAALLTAQIERPAPGFRDRIHVHSDVTNAEDLEAWPNL 418  
QY 454 IGGNVHVMESLDQMLWRPLPELSCHRVFGADGLYLTGASTHPGGVSGASGSAARIA 513  
DB 419 VGGDI-----TAGSALLRRPPTKIGECT-----YMASASNADPGGVHGMPCGWAAQAV 466  
QY 514 LSDSR 518  
|:|

b 467 LADHR 471

ESULT 10

RO22615

D AAO22615 standard; protein; 501 AA.

X X C AAO22615;

C X C AAO22615;

X T T 15-MAY-2003 (first entry)

X X X Synechocystis protein, SEQ ID No 20.

E Synechocystis

X X X Carotenoid isomerase catalytic activity; CRTISO; cis-carotenoid;

X X X all-trans geometric isomer; phytoene; phytofluene; zeta-carotene;

X X X neurosporene; lycopene; plant; dark; non-photosynthetic tissue; enzyme;

X X X synechocystis.

X X X Synechocystis sp.

S Synechocystis

X X X WO2003008534-A2.

X X X 30-JAN-2003.

X X X 18-JUL-2002; 2002WO-IL000600.

X X X 19-JUL-2001; 2001US-0306144P.

X X X (YISS ) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.

X X X Hirschberg J, Zamir D, Isaacson T;

X X X WPI; 2003-229567/22.

X X X Novel isolated polypeptide having carotenoids isomerase catalytic

X X X activity, useful for producing all-trans geometric isomers of cis-

X X X carotenoids, such as phytoene, phytofluene, zeta-carotene, neurosporene

X X X and lycopene.

X X X Disclosure; Page 134-135; 185pp; English.

X X X The invention relates to a novel isolated polypeptide having carotenoids

X X X isomerase catalytic activity, comprising a sequence at least 75 % similar

X X X to a 615 residue CRTISO amino acid sequence, given in the specification,

X X X as determined using the Standard protein-protein BLAST (blastp) software

X X X of the NCBI. The novel polypeptide is useful for producing the all-trans

X X X geometric isomers of cis-carotenoids, including phytoene, phytofluene,

X X X zeta-carotene, neurosporene and lycopene. The polypeptide enables

X X X carotenoid biosynthesis in plants in the dark and in non-photosynthetic

X X X tissues. This sequence represents a synechocystis protein relating to the

X X X carotenoid isomerase catalytic polypeptides of the invention

X X X Q Sequence 501 AA;

X X X Query Match

X X X Best Local Similarity 9.8%; Score 270; DB 6; Length 501;

X X X Matches 138; Conservative 78; Mismatches 224; Indels 128; Gaps 30;

X X X 1 MSAFDVAIVVGGHVALVSAAYLAKEGWSVEVLEKDTVLGGAVSTVERPFGKYVDRGSSA 60

X X X 3 VSPSYDAIVGSGIGLVTATQVSKGLKVLVLYLPFGSGAGYFER-EGYRFDVGNAS- 60

X X X 61 HLMIRHSGIIEELGHNGLRIYDC-----DPAFAPPAGTGGPIVPHR 106

X X X 61 --MI-----FGFGDRGTTNLLTRALAAVQALETLPVQIHVLPGLDKV--HR 108

X X X 107 DLDATCQSIERAGTKDADAVRRFV-AVWS-----ERSRVMKAFSTPTGTS 152

X X X 109 EYEAFLQES-IAKFPQAGQIRRFDECHQVNCNLTMTWELLSLEPRVLMVFQHP---- 164

X X X 153 NLIGAFPGSLA--TAGNSLSRQFTAPGDALLDEYFDSEALKAALAWFGAOSGPFMBEPG 210

X X X 165 ---GACLGKVLQPNVGVGIARHIDQPDLL--KFIDME---CYCW-----SVVPAD 208

QY 211 TAPMVGFAALM---HVLPPGRAVGGSGALSAAASRMVADGATVALGDGVTS-IRNSNH 266  
 DB 209 LTPMINAGWVFSDRHYGGINYPKGGVGGIABSLVAGLEKFGGKIRYGARVTKIIGNNOA 268  
 QY 267 WTVTTESREVEARKVIAGCHLLTLLDLNGGFFORT---TLDHWRKIRVPG-IGAVL 322  
 DB 269 IGVELANGEKIYGRIVNSA---TRWDTFGALTGDOPLPGKEKRRWRNRYQSSPFLSLHL 325  
 QY 323 RLATSALPSYRGDATTRESTSGQLLVSDRAHLRTAHGAALAGELPPRPVAVLGMFSFGID 382  
 DB 326 GVEADLLP-----EGTECHILLRDWDLEKEQGTIFVS-----IPTLL-----D 365  
 QY 383 PITAPAGRHQV-TLWSQW-QPYRLSGHEDWASVABAEADRIVGEGEAPAGFTDSVLDRF 440  
 DB 366 PSLAPDGYHIIHTFTPSWLESQNLSPQYEAKEGADSKLIDLEAIFPGL-DRALDYM 424  
 QY 441 -IQTPRDIESELGMIGGNVHVMSLDQMLWRPLPELSGHRVPG-----ADGLY 489  
 DB 425 EIGTFSHRFRFLGRQGT-----YGPFR---RRLPOLLPMFPENRTAIFGLY 468  
 QY 490 LTGASTHPGGVSGA--SGRSAA-RIAL 514  
 DB 469 CVGDSSTFPGQGLNAVAFSGFACAHRLAV 496

## RESULT 11

AAW22499

ID AAW22499 standard; protein; 582 AA.

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;

XX AC AAW22499;



489 YLTGASTHPOGGV-----SG--ASGRSAARIALS-----DSRRGKASQMMR 527  
561 YCVGDSOFGQGVIAVAFSGWMCARVAADLGPEKSDVLDLSALLRLGLWR 612

RESULT 13  
AB85728  
D AAB85728 standard; protein; 548 AA.

C AAB85728;  
X 29-OCT-2001 (first entry)  
T Enzyme involved in carotenoid biosynthetic pathway.  
E Carotenoid; enzyme; catalyst; geranylgeranylpyrophosphate, phytoene;  
W farnesylpyrophosphate.  
X Brevibacterium flavum.  
X JP2001149077-A.

D 05-JUN-2001.

F 24-NOV-1999; 99JP-00333395.

X 24-NOV-1999; 99JP-00333395.

X (MITU ) MITSUBISHI CHEM CORP.

A WPI; 2001-499379/55.

R N-PSDB; AAB76025.

X New DNA encoding enzyme of new carotenoid biosynthetic path, useful for

T preparation of carotenoid.

S Claim 5; Page 23-25; 33pp; Japanese.

X The invention relates to enzymes involved in a carotenoid biosynthetic  
C pathway and DNA sequences encoding the polypeptides. One polypeptide has  
C the activity of catalyzing a reaction forming geranylgeranylpyrophosphate  
C from farnesylpyrophosphate. Another polypeptide has an activity of  
C catalyzing a reaction forming phytoene from geranylgeranylpyrophosphate.  
C The DNA can be used for the preparation of carotenoid. The present  
C sequence represents an enzyme of carotenoid biosynthetic pathway

X Sequence 548 AA;

Query Match 8.0%; Score 221.5; DB 4; Length 548;  
Best Local Similarity 25.8%; Pred. No. 1.1e-09;  
Matches 141; Conservative 66; Mismatches 236; Indels 103; Gaps 27;

Y 7 AVVVGSHNALVSAAYLAREGNSVELEKTVLGGAVTVE--RPPGYKVDKGSAAHLM 64  
b 14 AVVVGAGVAGLATSALLARDGQVTVLEKTVGGRAGSLSEIDPGRWDTGFWILMP 73  
Y 65 RSGIIEELGLAHGURYIDCFMFAPAPPAPGDTGGIVF---HRDL-----ATC 112  
b 74 EAFDHPFAL-FCARTSDYLD-----LVELTPGYR----VFGTHDAVDVPGREKAIALF 123  
Y 113 QSTERACGTK-----DADAYRRFVAVNSRSRHVMKAFST--PPTGSLNIGAFGGIAT 163  
b 124 ESEPGAGAKLGNLYDSADAYD--IAI-----DRFLYNNFSTLGPLLHRDVLTRAGRLF- 176  
Y 164 ARGNSLSQFLAPGDALLDEYDSEALKAAALWFGAGSGPPM---SEPGTAPMVGPAAL 220  
b 177 ----SLTTRSL-----QKYVNSQFSSPVLKQLTY-----PAVFLSSRPTTTPSM-VHLM 221  
Y 221 MHV-LPPG--RAVGGGALSALASRMVADGATVAGDVTGVSIRENSNHWITVTTESGREV 277  
b 222 SHFDLVQGVKYPGGGTAVVNAHQLALENGVEFDSEVISINTASSRGNTSATGVSFL 281  
Y 278 HARKVIAGCHIITLTD---LLNGGDFDRTTLDHWRKIRV-----GPGIGAVLRL- 324

Db 282 HNRKV-----QNLDAVLVWSAGDLHHTENNLLPRELATYPERYWSNRPGLGAVLILL 334  
Qy 325 -ATSAIPSY-RGDAITRETSGLQLLVSDRAHLRTAKGAALAGELPPRPAVLGMSFGID 382  
Db 335 GVKGELPQLDHHNLFFESDWTDDFAVVDPGQPTRPHNAS-----NSIYVSKFSTSE 386  
Qy 383 PTIAPAGRHOVTLWSQWQFYRLSGHRD-WASVAEABADRIVGE-----MEAPAPGFTDS 435  
Db 387 DGVAPAGYENLFLIIPAKASSISIGHGDAYWQASASVETIASHAINQIATQAGIDPLTDR 446  
Qy 436 VLDRFTQTPRDIASELGMIGNVHVMESLDQWMLWRPLPELSGHRVPGADGGLYLTGAST 495  
Db 447 IVVKRTIGPADFHRYHSWVSGALGPAHTLROSAFLR--GRNSSRKV----DNLFYSGATT 501  
Qy 496 HPGGGV 501  
Db 502 VPGVGI 507

RESULT 14

AAB76640

ID AAB76640 standard; protein; 548 AA.

XX AAB76640;

XX 11-APR-2001 (first entry)

XX Corynebacterium glutamicum MCT protein SEQ ID NO:262.

XX Corynebacterium glutamicum; Brevibacterium lactofermentum; MCT;  
XX membrane construction and membrane transport protein; petroleum spill;  
XX hydrocarbon degradation; gram positive aerobic bacterium; marker;  
XX identification; microorganism; fine chemical production; transformation;  
XX genome mapping; genetic engineering.

XX Corynebacterium glutamicum.

XX WO200100805-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000MO-IB000926.

XX 25-JUN-1999; 99US-0141031P.  
PR 08-JUL-1999; 99DE-01031454.  
PR 08-JUL-1999; 99DE-01031478.  
PR 08-JUL-1999; 99DE-01031563.  
PR 08-JUL-1999; 99DE-01032122.  
PR 08-JUL-1999; 99DE-01032124.  
PR 08-JUL-1999; 99DE-01032125.  
PR 08-JUL-1999; 99DE-01032128.  
PR 08-JUL-1999; 99DE-01032180.  
PR 08-JUL-1999; 99DE-01032182.  
PR 08-JUL-1999; 99DE-01032190.  
PR 08-JUL-1999; 99DE-01032191.  
PR 08-JUL-1999; 99DE-01032209.  
PR 08-JUL-1999; 99DE-01032212.  
PR 08-JUL-1999; 99DE-01032227.  
PR 08-JUL-1999; 99DE-01032228.  
PR 08-JUL-1999; 99DE-01032229.  
PR 08-JUL-1999; 99DE-01032230.  
PR 14-JUL-1999; 99DE-01032927.  
PR 14-JUL-1999; 99DE-01033005.  
PR 14-JUL-1999; 99DE-01033006.  
PR 27-AUG-1999; 99DE-01040764.  
PR 27-AUG-1999; 99DE-01040765.  
PR 27-AUG-1999; 99DE-01040766.  
PR 27-AUG-1999; 99DE-01040830.  
PR 27-AUG-1999; 99DE-01040831.  
PR 27-AUG-1999; 99DE-01040832.  
PR 27-AUG-1999; 99DE-01040833.  
PR 31-AUG-1999; 99DE-01041378.



31-AUG-1999; 99DE-01041379.  
31-AUG-1999; 99DE-01041395.  
03-SEP-1999; 99DE-01042077.  
03-SEP-1999; 99DE-01042078.  
03-SEP-1999; 99DE-01042079.  
03-SEP-1999; 99DE-01042088.

(BADI ) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;

WPI; 2001-071486/08.  
N-PSDB; AAF67873.

Corynebacterium glutamicum nucleic acids encoding membrane construction and membrane transport proteins or their portions, useful for typing or identifying C. glutamicum or related bacteria, and as markers for transformation.

Claim 20; Page 534-536; 1119pp; English.

AA67743 to AA68080 encode the Corynebacterium glutamicum membrane construction and membrane transport (MCT) proteins given in AAB76510 to AAB76847. The MCT nucleic acids and proteins are useful in the identification of microorganisms which can be used to produce fine chemicals, for modulating fine chemical production in C. glutamicum or related bacteria (e.g. Brevibacterium lactofermentum), the typing or identification of C. glutamicum or related bacteria, as reference points for mapping C. glutamicum genome, and as markers for transformation. AA68082 and AA68082 represent sequencing primers which are used in an example from the present invention

Sequence 548 AA;

Query Match 7.98; Score 219.5; DB 4; Length 548;  
Best Local Similarity 25.88; Pred. No. 1.6e-09;  
Matches 139; Conservative 68; Mismatches 243; Indels 89; Gaps 26;  
N-PSDB; AAF67873.  
7 AVVVGSHALVSAAYLAREGWSVELEKDTVLGGAVSTVE--RFGYKVDGSSAHMI 64  
14 AVVIGAGVAGLATSALLARDGWQVLEKTDVGRAGSLISGFFGPRNDTGPWYLP 73  
65 RHSGIIBELGANGLYID---CDPW--AFAPAPGDTGPGIVFHRDLDAQCQSIERAC 119  
74 EAFDHPFAL-FGACTSYLDLVELTGPYRFGSGTHDAVDVP--TGEEAIALFESTIEPCA 130  
120 GTK-----DADAYRFRVAVWSRSRHVKAFST--PPTGSLNIGAFGLATARGNSL 170  
131 GAKLNVLDSDAYD--IAL----DRFLVNFSTLGLLHRDVLTRAGRLP-----SL 179  
171 SRQFLAFCDALLDEYFDSALKALAWFGAQSGPPM---SEPCAPMVGFALMHV-LPP 226  
180 TRSL---QKVNQSFSSPVLRLITY-----PAVFLSSRPPTTFSM-YHLSMHTDLVQ 228  
227 G--RAVGGSGALSAALASRAVDGATVALGDGVTISIRNSNHWVTTESGREVHARKVIA 284  
229 GVKYPIGFTAVNHALHQLALENVEFQDSEVISINTASSRGNTSATGVSLHNNKRV-- 286  
285 GCHLTITLD---LIGNCGGFTITLDHWRRIKRV-----GFGIGAVLRL--AFSALP 330  
287 -----QNLADLVVSAGDLHHTENLLPRELTYPERYNSNRNFGIGAVILLGVKGELP 341  
331 SY-RGDATTTESTSGQLLYSDRAHLRTANGALAGELPPRAVLGMSFGSDPTTAPAG 389  
342 QLDHNNLFFESDWDTDFAVVFQDGLTRPNAS-----NSIYVSKPSTSEDGVAPAG 393  
390 RHQVTLWSQMPYRLSGHRD-WASVAEAEADRIVG8-----MEAFAPGFTSVLDRFTQ 442  
394 YENLVLPYTKASSIGCHDAYQWSASASVETASHAINQIAQAGIPDLTDEIVVKRYI 453  
443 TPRDIESLGMIGNVHMVENSMDQMLWRLPELSCHRVPGADGLYLTCASHTPGGV 501  
454 GPADFEHRYHSWVSGALGPAHTLRQSAFLR--GRNSSRKV---NNIFYSGATTVPVGI 507

RESULT 15  
AAB76641

ID AAB76641 standard; protein; 548 AA.

XX AAB76641;

XX 11-APR-2001 (first entry)

Corynebacterium glutamicum MCT protein SEQ ID NO:264.

Corynebacterium glutamicum; Brevibacterium lactofermentum; MCT;  
membrane construction and membrane transport protein; petroleum spill;  
hydrocarbon degradation; gram positive aerobic bacterium; marker;  
KW identification; microorganism; fine chemical production; transformation;  
genome mapping; genetic engineering.

OS Corynebacterium glutamicum.

XX WO200100805-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB000926.

XX 25-JUN-1999; 99US-0141031P.  
XX 08-JUL-1999; 99DE-01031454.  
XX 08-JUL-1999; 99DE-01031478.  
XX 08-JUL-1999; 99DE-01031563.  
XX 09-JUL-1999; 99DE-01032122.  
XX 09-JUL-1999; 99DE-01032124.  
XX 09-JUL-1999; 99DE-01032125.  
XX 09-JUL-1999; 99DE-01032128.  
XX 09-JUL-1999; 99DE-01032180.  
XX 09-JUL-1999; 99DE-01032182.  
XX 09-JUL-1999; 99DE-01032190.  
XX 09-JUL-1999; 99DE-01032191.  
XX 09-JUL-1999; 99DE-01032209.  
XX 09-JUL-1999; 99DE-01032212.  
XX 09-JUL-1999; 99DE-01032227.  
XX 09-JUL-1999; 99DE-01032228.  
XX 09-JUL-1999; 99DE-01032229.  
XX 09-JUL-1999; 99DE-01032230.  
XX 14-JUL-1999; 99DE-01032927.  
XX 14-JUL-1999; 99DE-01033005.  
XX 14-JUL-1999; 99DE-01033006.  
XX 27-AUG-1999; 99DE-01040764.  
XX 27-AUG-1999; 99DE-01040765.  
XX 27-AUG-1999; 99DE-01040766.  
XX 27-AUG-1999; 99DE-01040830.  
XX 27-AUG-1999; 99DE-01040831.  
XX 27-AUG-1999; 99DE-01040832.  
XX 27-AUG-1999; 99DE-01040833.  
XX 31-AUG-1999; 99DE-01041378.  
XX 31-AUG-1999; 99DE-01041379.  
XX 31-AUG-1999; 99DE-01041395.  
XX 03-SEP-1999; 99DE-01042077.  
XX 03-SEP-1999; 99DE-01042078.  
XX 03-SEP-1999; 99DE-01042079.  
XX 03-SEP-1999; 99DE-01042088.

(BADI ) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G;

WPI; 2001-071486/08.  
N-PSDB; AAF67874.

Corynebacterium glutamicum nucleic acids encoding membrane construction and membrane transport proteins or their portions, useful for typing or identifying C. glutamicum or related bacteria, and as markers for transformation.

```
x S Claim 20; Page 538-540; 1119pp; English.
x C AAF57743 to AAF68080 encode the Corynebacterium glutamicum membrane
x C construction and membrane transport (MCT) proteins given in AAF76510 to
x C AAF76847. The MCT nucleic acids and proteins are useful in the
x C identification of microorganisms which can be used to produce fine
x C chemicals, for modulating fine chemical production in C. glutamicum or
x C related bacteria (e.g. Brevibacterium lactofermentum), the typing or
x C identification of C. glutamicum or related bacteria, as reference points
x C for mapping C. glutamicum genome, and as markers for transformation.
x C AAF68082 and AAF68082 represent sequencing primers which are used in an
x C example from the present invention
x Q Sequence 548 AA;
Query Match 7.9%; Score 219.5; DB 4; Length 548;
Best Local Similarity 25.9%; Pred. No. 1.e-09;
Matches 139; Conservative 68; Mismatches 243; Indels 89; Gaps 26;
7 AVVVGSGHVALVSAAYLAREGWSVEVLEKTVLGGAVSTVE--RPGYKYVDRGSSAHIMI 64
14 AVVIGAGVAGLATSALLARDGMQVLEKNTDVGGRAGSLEISGPGFRMDTGPSWYIMP 73
65 RMSGIIEELGAGHLYID--CDPW--AFAPPAPGTDGPGIVFHRDLDTATCOTIERAC 119
74 EAFDHFAL--FGACTSDVLDVLELTPGYRVFSGTHDAVDP--TCREAIALPESIEPGA 130
120 GTK-----DADAYRRFVAVWSESRHVMKAPST--PPTGNSNLIGAPGGLATARGNSL 170
131 GAKLGNLYLSAADAVID--IAI----DRFLYNNFSTGLPLLRDVLTRAGRLF-----SL 179
171 SRQFLAPGDALLDEYFDSBALKAALWFGAQSGPPM---SEPCTAPMVGFAALMHV-LPP 226
180 TRSL-----QKYVNSQFSSPVLQILTY-----PAVFLSSRPPTTTPSM-YHLMSTDLVQ 228
227 G--RAVGGSGALSAAALASRMAYDGAIVAGDGVTISIRNSNHWITVTSSGREVHARKYIA 284
229 GVKYPGGFTAVVNALHQLALENGVEFQIDSEVISINTASSRGNTSANGVELLHNRKV-- 286
285 GCHILTLTLD---LLGNGGFDRTLDHNRKIV-----GPGIGAVLEL--ATSALP 330
287 -----QNLDAIVVSGADLHHTENLLPRELTYPERYWSNRNPGIGAVLILLGVKGELP 341
331 SY-RGDATTRESTSGILIVSDRAHLRTAHGAALAGELPPRPVAVLZMSFSGIDPTIAPAG 389
342 QLDHNLFTSEDTDDFAVVPDGPQLTRPHNAS-----NSIYVSKPSTSEDGVAPAG 393
390 RHQVTLMSQWQPYRLSGHRD--WASVAEAEADRIUGE-----MEAFPGFTDVLDRFIQ 442
394 YENLFVLIPTKASSSTIGHGDATYMQASASVETIASHAINQIATQAGIPDLTDRIIVKRTI 453
443 TPRDIESELGMIGGVNMHVEMSLDQMLWRPLPELSGHRVPCADGLYLTGASTHPGGGV 501
454 GPADFPHRYHSWVGALGPAHTLRQSAFLR--GRNSRKV---NNLFYSGATTVGVGI 507
```

Search completed: February 29, 2004, 14:44:12  
Ob time : 66.9999 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

{ protein - protein search, using sw model

in on: February 29, 2004, 14:33:49 ; Search time 14.3897 Seconds  
(without alignments)  
3837.172 Million cell updates/sec

itle: US-09-941-947a-36

affect score: 956  
sequence: 1 MLWLNALIVFTVVGMEW.....ARSGAARDQGVDTSSSGK 175

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

1: SP\_TREMBL 25.\*

2: SP\_archaea.\*

3: SP\_bacteria.\*

4: SP\_fungi.\*

5: SP\_human.\*

6: SP\_invertebrate.\*

7: SP\_mammal.\*

8: SP\_mhc.\*

9: SP\_organelle.\*

10: SP\_page.\*

11: SP\_plant.\*

12: SP\_rodent.\*

13: SP\_virus.\*

14: SP\_vertibrate.\*

15: SP\_unclassified.\*

16: SP\_rvirus.\*

17: SP\_bacteriap.\*

17: SP\_archheap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	956	100.0	175	2 Q8GCR9	Q8GCR9 pantoea ste
2	822	86.0	175	2 Q8VUJ3	Q8VUJ3 pantoea agg
3	492	51.5	162	2 Q8RLH6	Q8RLH6 paracoccus
4	487	50.9	169	2 P94792	P94792 flavobacter
5	468	49.0	155	16 Q88HV7	Q88HV7 pseudomonas
6	271	28.3	151	17 Q97UT7	Q97UT7 sulfobolus
7	271	28.3	303	10 Q9LTG0	Q9LTG0 arabidopsi
8	270	28.2	308	10 Q9FV13	Q9FV13 narcissus p
9	270	28.2	322	10 Q9SPK6	Q9SPK6 haematococ
10	263.5	27.6	311	10 Q9FTN2	Q9FTN2 citrus unsh
11	263.5	27.6	311	10 Q9FUB9	Q9FUB9 citrus unsh
12	263	27.5	299	10 Q8LVK3	Q8LVK3 vitis vinif
13	262.5	27.5	305	10 Q8VXP2	Q8VXP2 crocus sati
14	261	27.3	315	10 Q49815	Q49815 capsicum an
15	260.5	27.2	316	10 Q49814	Q49814 capsicum an
16	257.5	26.9	314	10 Q986Y0	Q986Y0 lycopersico

17	257	26.9	296	10	Q84U11	Q84U11 haematococ
18	256.5	26.8	148	2	Q7WT73	Q7WT73 marine bact
19	254	26.6	309	10	Q986Y1	Q986Y1 lycopersico
20	253	26.5	294	10	Q96297	Q96297 arabidopsi
21	253	26.5	309	10	Q9FV41	Q9FV41 tegetes ere
22	253	26.5	310	10	Q9S228	Q9S228 arabidopsi
23	247	25.8	292	10	Q8LN34	Q8LN34 oryza sativ
24	245	25.6	309	10	Q7XQ82	Q7XQ82 oryza sativ
25	231	24.2	127	10	Q8RY53	Q8RY53 sandersonia
26	98.5	10.3	410	16	P74167	P74167 synechocyst
27	96.5	10.1	324	16	Q8XZ21	Q8XZ21 raietonia s
28	96	10.0	315	16	Q98MP4	Q98MP4 rhizobium l
29	93.5	9.8	355	16	Q986V0	Q986V0 rhizobium l
30	93	9.7	418	10	Q8W3L3	Q8W3L3 chlorella v
31	92	9.6	408	2	Q8GCR29	Q8GCR29 streptomyc
32	91.5	9.6	391	10	Q9ZPP7	Q9ZPP7 perilla fru
33	91	9.5	313	16	Q8UIH7	Q8UIH7 agrobacteri
34	90.5	9.5	444	10	Q8HLX8	Q8HLX8 betula verr
35	89	9.3	163	10	Q9SX07	Q9SX07 arabidopsi
36	88	9.2	301	16	Q9ISM4	Q9ISM4 pseudomonas
37	87.5	9.2	302	10	Q93Y93	Q93Y93 brassica ra
38	87	9.1	335	16	Q8ZG25	Q8ZG25 streptomyc
39	87	9.1	600	5	Q961U9	Q961U9 drosophila
40	87	9.1	674	5	Q9VIK2	Q9VIK2 drosophila
41	86.5	9.0	104	10	Q9XH55	Q9XH55 brassica na
42	86.5	9.0	108	10	Q9XH56	Q9XH56 brassica ol
43	86.5	9.0	429	10	Q9M4D4	Q9M4D4 brassica ju
44	86.5	9.0	435	10	Q8W4M9	Q8W4M9 arabidopsi
45	86	9.0	445	16	Q8PQ16	Q8PQ16 xanthomonas

#### ALIGNMENTS

#### RESULT 1

Q8GCR9 PRELIMINARY; PRT; 175 AA.  
AC Q8GCR9  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Beta-carotene hydroxylase.  
GN CRTZ.  
OS Pantoea Stewartii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pantoea.  
OX NCBI\_TaxID=66269;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 8200;  
RA deSouza M.L., Kollmann S.R., Schroeder W.A.;  
RT "Carotenoid Biosynthesis (WO 02/079395 A2)";  
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY166713; AAN85601.1; -  
DR GO; GO:0003824; F:catalytic activity; IEA.  
DR GO; GO:0008152; P:metabolism; IEA.  
DR InterPro; IPR005596; Carotene hydrox.  
DR InterPro; IPR006087; Sterol desat.  
DR Pfam; PF03897; Carotene hydrox.; 1.  
SQ SEQUENCE 175 AA; 19884 MW; D7F1069562D83054 CRC64;

Query Match 100.0%; Score 956; DB 2; Length 175;

Best Local Similarity 100.0%; Pred. No. 4.8e-89;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MLWLNALIVFTVVGMEVVAALAHKYLIMHGWGNGWLSHHEPRKGAFVNDLVAVVFAI	60
Db	1	MLWLNALIVFTVVGMEVVAALAHKYLIMHGWGNGWLSHHEPRKGAFVNDLVAVVFAI	60
QY	61	VSIALIYFGSTGIMPLQWIGAGMTAYGLLYFMVHDGLVHQWPPRYIPRKGYLEKLYNAH	120
Db	61	VSIALIYFGSTGIMPLQWIGAGMTAYGLLYFMVHDGLVHQWPPRYIPRKGYLEKLYNAH	120

121 RMHVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDEQGVDTSSGK 175  
121 RMHVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDEQGVDTSSGK 175

## RESULT 2

Q8VUJ3 PRELIMINARY; PRT; 175 AA.  
Q8VUJ3, 2002 (TREMELrel. 20, Created)  
01-MAY-2000 (TREMELrel. 23, Last sequence update)  
01-OCT-2003 (TREMELrel. 25, Last annotation update)  
CrtZ protein.  
N CRTZ  
S Pantoea agglomerans pv. milletiae.  
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
C Enterobacteriaceae; Pantoea.  
X NCBI\_TaxID=182454;  
N N  
P SEQUENCE FROM N.A.  
A Kamiunten H., Hirata R.;  
T "Isolation and characterization of carotenoid biosynthesis genes from  
T Pantoea agglomerans pv. milletiae Wist 801.";  
L Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
R EMBL; AB076662; BAB79605.1; -  
R GO: GO:0003824; P: catalytic activity; IEA.  
R GO: GO:0008152; P: metabolism; IEA.  
R InterPro; IPR005596; Carotene\_hydrox.  
R InterPro; IPR006087; Sterol\_desat.  
R Pfam; PF03897; Carotene\_hydrox; 1.  
Q SEQUENCE 175 AA; 20222 MW; 4849DD4C61167845 CRC64;

Query Match 86.0%; Score 822; DB 2; Length 175;  
Best Local Similarity 82.3%; Pred. No. 1.9e-75;  
Matches 144; Conservative 13; Mismatches 18; Indels 0; Gaps 0;

1 MNTWALIVFTVVGMEVVAALAHKYIMHG-WGHWLSSH-EPKGAPEVNDLYAVVFAI 50  
1 MNTWALIVFTVVGMEVVAALAHKYIMHG-WGHWLSSH-EPKGAPEVNDLYAVVFAA 60  
61 VSIALYFGSTGIWPLQWICAGMTAYGLLYFMVHDLVHQRWPFYIPRKGYLRLYMAH 120  
61 LSILLIYVAGSTGWPQLWICAGMTAYGLLYFMVHDLVHQRWPFYIPRKGYLRLYMAH 120  
121 RMHVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDEQGVDTSSGK 175  
121 RMHVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDEQGVDTSSGK 175

## RESULT 3

Q8RLH6 PRELIMINARY; PRT; 162 AA.  
Q8RLH6, 2000 (TREMELrel. 13, Created)  
01-MAY-2000 (TREMELrel. 13, Last sequence update)  
01-OCT-2003 (TREMELrel. 25, Last annotation update)  
Carotene hydroxylase.  
N CRTZ  
S Paracoccus marcusii.  
C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;  
C Rhodobacteraceae; Paracoccus.  
X NCBI\_TaxID=59779;  
N N  
P SEQUENCE FROM N.A.  
A Harker M., Hirschberg J.;  
T "Carotenoid biosynthesis genes in the bacterium Paracoccus marcusii  
T Hml.";  
L Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
R EMBL; Y15112; CAB56060.1; -  
R GO: GO:0003824; P: catalytic activity; IEA.  
R GO: GO:0008152; P: metabolism; IEA.  
R InterPro; IPR005596; Carotene\_hydrox.

DR InterPro; IPR006087; Sterol\_desat.  
DR Pfam; PF03897; Carotene\_hydrox; 1.  
SQ SEQUENCE 162 AA; 18882 MW; 798891BFF8743999 CRC64;

Query Match 51.5%; Score 492; DB 2; Length 162;  
Best Local Similarity 57.4%; Pred. No. 5.3e-42;  
Matches 93; Conservative 20; Mismatches 39; Indels 10; Gaps 4;

6 NALIVFTVVGMEVVAALAHKYIMHG-WGHWLSSH-EPKGAPEVNDLYAVVFAIVSIA 64  
3 NFLIVATVLMELTAYSVRHWIMHGFLGWGWHKSHHEDHLEKNDLYGLVFAIVATV 62  
65 LIYFGSTGIWPLQWICAGMTAYGLLYFMVHDLVHQRWPFYIPRKGYLRLYMAHRM 122  
63 LFTVG-WIAPVLMIALGMTVYGLIYFVLDGVHQRWPFYIPRKGYARLLYQAHRL 120  
123 HHAVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDE 164  
121 HHAVEGRDHCVSFGFTYAPVDKLDK-----TSGVLRAE 157

## RESULT 4

P94792 PRELIMINARY; PRT; 169 AA.  
AC P94792;  
DT 01-MAY-1997 (TREMELrel. 03, Created)  
DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DN B-carotene hydroxylase.  
GN CRTZ  
OS Flavobacterium sp. ATCC 21588.  
OC Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales;  
OC Flavobacteriaceae; Flavobacterium.  
OX NCBI\_TaxID=50286;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RL534;  
RX MEDLINE=97186694; PubMed=9034310;  
RA Pasamonles L., Hug D., Tessier M., Hohmann H.P., Schierle J.,  
RA van Loon A.P.;  
RT "Isolation and characterization of the carotenoid biosynthesis genes  
RT of Flavobacterium sp. strain RL534.";  
RL Gene 185:35-41(1997).  
DR EMBL; U62808; AAC44852.1; -  
DR GO: GO:0005509; P: calcium ion binding; IEA.  
DR GO: GO:0003824; P: catalytic activity; IEA.  
DR GO: GO:0008152; P: metabolism; IEA.  
DR InterPro; IPR005596; Carotene\_hydrox.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR006087; Sterol\_desat.  
DR Pfam; PF03897; Carotene\_hydrox; 1.  
DR PROSITE; PS00018; EF\_HAND; 1.  
SQ SEQUENCE 169 AA; 19282 MW; 85334038DB3DA81 CRC64;

Query Match 50.9%; Score 487; DB 2; Length 169;  
Best Local Similarity 51.5%; Pred. No. 1.8e-41;  
Matches 88; Conservative 31; Mismatches 44; Indels 8; Gaps 4;

5 WNALI-VFTVVGMEVVAALAHKYIMHG-WGHWLSSH-EPKGAPEVNDLYAVVFAIV 61  
4 WAAILYVLTVAAMELTAYSVRHWIMHGFLGWGWHKSHHEDHLEKNDLYGLVFAIV 63  
62 SIALLYFGSTGIWPLQWICAGMTAYGLLYFMVHDLVHQRWPFYIPRKGYLRLYMAHR 121  
64 SIVLPAIGAGSDLAWLAVGTCYGLIYFLHDLGVHQRWPFYIPRKGYLRLYMAHR 123  
122 MHAVRGKGGCVSFGFLYAPPLSKLQATLRERHARSGAARDEQGVDTSS 172  
124 MHAVHAGRENCVSPGFINAPSVDLSKELK-----RSGALLKOREGADNRT 169

## RESULT 5

Q88HV7



```
RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RA Arabidopsis Open Reading Frame (ORF) Clones";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AB025606; BAA98075.1; -
DR EMBL; AY074394; AAL67090.1; -
DR EMBL; AY117225; AAM51300.1; -
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005596; Carotene_hydrox.
DR InterPro; IPR006087; Sterol_desat.
DR Pfam; PF03897; Carotene_hydrox; 1.
SQ SEQUENCE 303 AA; 33777 MW; C7B8C9BF8BAE34CB CRC64;

Query Match 28.3%; Score 271; DB 10; Length 303;
Best Local Similarity 42.0%; Pred. No. 2.9e-19;
Matches 55; Conservative 26; Mismatches 44; Indels 6; Gaps 3;

2y 15 VGMEVVAALAHKYIMHGWGWSHSHHEPRKGAPEVNDLYAVVFAIVSIALIYFG--STG 72
2b 145 VGMEFWARWAHRAHWDLSNWNCHESHKHPREGAPELNDVFAITNAVPAIGLGYGFIKNG 204
2y 73 IWPLQWIGAGM--TAYGLLFVMDHGLVHQRWPRYIPRKGYLKRLYMAHRMHAHVRGKE 130
2b 205 LVPLGLCFGAGLGTIFMGCAVMFVHDGLVHKRFPVGPPIANVPYILRKVAARAHLEHTDKPKG 264
2y 131 GCVSFGFLYAP 141
2b 265 --VPYGLFLGP 273

RESULT 8
2y Q9FY13 PRELIMINARY; PRT; 308 AA.
2b Q9FY13;
2y 01-MAR-2001 (TREMBlrel. 16, Created)
2b 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
2y 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
2b 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
2y Beta-carotene hydroxylase.
2b Narcissus pseudonarcissus (Daffodil).
2y Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
2b Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
2y Narcissus.
2b NCBI_TaxID=39639;
2y [1]
2b TISSUE=Flower;
2y Schaub P., Beyer P., Al-Babili S.;
2b "A cDNA encoding beta-carotene hydroxylase from Narcissus
pseudonarcissus.";
2y Submitted (AUG-2000) to the EMBL/GenBank/DBSJ databases.
2b EMBL; AJ278882; CAC06712.1; -
2y GO; GO:0003824; F:catalytic activity; IEA.
2b GO; GO:0008152; P:metabolism; IEA.
2y InterPro; IPR005596; Carotene_hydrox.
2y InterPro; IPR006087; Sterol_desat.
2y Pfam; PF03897; Carotene_hydrox; 1.
SQ SEQUENCE 308 AA; 34849 MW; 2782E3027252D30C CRC64;

Query Match 28.2%; Score 270; DB 10; Length 308;
Best Local Similarity 43.5%; Pred. No. 3.7e-19;
Matches 57; Conservative 22; Mismatches 46; Indels 6; Gaps 3;

2y 15 VGMEVVAALAHKYIMHGWGWSHSHHEPRKGAPEVNDLYAVVFAIVSIALIYFG--STG 72
2b 150 VGMEFWARWAHRAHWSLWMEHSHKPRDGPFLNDVFAVNAVPAISLLIYGFNKG 209
2y 73 IWPLQWIGAGM--TAYGLLFVMDHGLVHQRWPRYIPRKGYLKRLYMAHRMHAHVRGKE 130
2b 210 LVPLGLCFGAGLGTIFMGCAVMFVHDGLVHRRFPVGPPIANVPYILRKVAARAHLEHTDKPKG 267
2y 131 GCVSFGFLYAP 141
```

```
Db 268 NGVPYGLFLGP 278

RESULT 9
Q9SPK6 PRELIMINARY; PRT; 322 AA.
ID Q9SPK6
AC Q9SPK6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Carotenoid hydroxylase (Fragment).
OS Haematococcus pluvialis.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Haematococaceae; Haematococcus.
OX NCBI_TaxID=44745;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99453720; PubMed=10524195;
RA Linden H.;
RT "Carotenoid hydroxylase from Haematococcus pluvialis: cDNA sequence,
regulation and functional complementation.";
RL Biochim. Biophys. Acta 1446:203-212 (1999).
DR EMBL; AF162276; AAD54243.1; -
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005596; Carotene_hydrox.
DR InterPro; IPR006087; Sterol_desat.
DR Pfam; PF03897; Carotene_hydrox; 1.
DR PROSITE; PS00626; RCC1_2; 1.
FT NON TER 1
SQ SEQUENCE 322 AA; 34986 MW; 45100BA5ECBA13EC CRC64;

Query Match 28.2%; Score 270; DB 10; Length 322;
Best Local Similarity 42.0%; Pred. No. 3.9e-19;
Matches 60; Conservative 20; Mismatches 49; Indels 14; Gaps 4;

Qy 8 LIVFTVTVGVMEVVAALAHKYIMHGWGWSHSHHEPRKGAPEVNDLYAVVFAIVSIAL 65
160 LLVVGGLGALGEMVARYAHKAINHESPLGWLHSHHETPTGPEENDLPAINGLPAMLL 219
Qy 66 IYFGSTGIWPLQWIGA-----GMTAYGLLFVMDHGLVHQRWPRYIPRKGYLKRLYM 118
220 CTF---GFPLNVLGACFGAGLGTIFMGCAVMFVHDGLVHRRFPVGPPIAGLEPYMKRLTV 276
Qy 119 AHRMHAHVRGKGCVSFGFLYAP 141
277 ARLHRS--GRYGGAPWGMFLGP 297

RESULT 10
Q9FUN2 PRELIMINARY; PRT; 311 AA.
ID Q9FUN2
AC Q9FUN2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Beta-carotene hydroxylase.
GN CHX1.
OS Citrus unshiu (Satsuma orange).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OX eurosids II; Sapindales; Rutaceae; Citrus.
OX NCBI_TaxID=55188;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Miyagawa; TISSUE=Fruit;
RA Kim I., Ko K., Kim C., Chung W.;
RT "Isolation of a cDNA encoding beta-carotene hydroxylase from Citrus.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF296158; AAG10793.1; -
DR GO; GO:0003824; F:catalytic activity; IEA.
```

```
R GO: GO:0008152; P:metabolism; IEA.
R InterPro; IPR005596; Carotene hydrox.
R InterPro; IPR006087; Sterol desat.
R Pfam; PF03897; Carotene hydrox. 1.
2 SEQUENCE 311 AA; 34759 MW; 37083C01A303D4A4 CRC64;

Query Match 27.6%; Score 263.5; DB 10; Length 311;
Best Local Similarity 39.3%; Pred. No. 1.7e-18;
Matches 59; Conservative 26; Mismatches 52; Indels 13; Gaps 4;

15 VGMVEVAALAHKYMHGWMGWSHHEPRKGAFVNDLYAVVFAIVSIALIYFG--STG 72
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 VGMVEFWARWAKALWHSLSHHEPRKGAFVNDLYAVVFAIVSIALIYFG 214
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

73 IWPLOWIGAGM--TAYGLLYFMVHDGLVHQWPPRYIPRKGYLKRLYMAHMHAVRGKE 130
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215 LVPLGCLFCGAGLGITVFGMAYNFVHDGLVHQWPPRYIPRKGYLKRLYMAHMHAVRGKE 272
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

131 GCVSFGFLYAPP-----LSKQATLRER 153
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 HGVPGYGLFGPKLEEVGGLLEKEISKR 302
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

SUITE 11
FUJ99
Q9FUB9 PRELIMINARY; PRT; 311 AA.
Q9FUB9
01-MAR-2001 (TRENBLrel. 16, Created)
01-MAR-2001 (TRENBLrel. 16, Last sequence update)
01-OCT-2003 (TRENBLrel. 25, Last annotation update)
Beta-carotene hydroxylase.
CHK2.
Citrus unshiu (Satsuma orange).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
eurosids II; Sapindales; Rutaceae; Citrus.
NCBI_TaxID=55188;
[1]
SEQUENCE FROM N.A.
STRAIN=cv. Miyagawa;
Kim I.-J., Ko K.-C., Kim C.-S., Chung W.-I.;
"Isolation of cDNA showing polymorphism to GenBank Accession Number
AF296158 from Citrus.";
Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF315289; AAG33636.1; -.
GO: GO:0003824; P:catalytic activity; IEA.
GO: GO:0008152; P:metabolism; IEA.
InterPro; IPR005596; Carotene hydrox.
InterPro; IPR006087; Sterol desat.
Pfam; PF03897; Carotene hydrox. 1.
SEQUENCE 311 AA; 34766 MW; E7868067F4A00AEB CRC64;

Query Match 27.6%; Score 263.5; DB 10; Length 311;
Best Local Similarity 39.3%; Pred. No. 1.7e-18;
Matches 59; Conservative 26; Mismatches 52; Indels 13; Gaps 4;

15 VGMVEVAALAHKYMHGWMGWSHHEPRKGAFVNDLYAVVFAIVSIALIYFG--STG 72
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 VGMVEFWARWAKALWHSLSHHEPRKGAFVNDLYAVVFAIVSIALIYFG 214
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

73 IWPLOWIGAGM--TAYGLLYFMVHDGLVHQWPPRYIPRKGYLKRLYMAHMHAVRGKE 130
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215 LVPLGCLFCGAGLGITVFGMAYNFVHDGLVHQWPPRYIPRKGYLKRLYMAHMHAVRGKE 272
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

131 GCVSFGFLYAPP-----LSKQATLRER 153
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
273 HGVPGYGLFGPKLEEVGGLLEKEISKR 302
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

SUITE 12
LKV3
Q8LKV3 PRELIMINARY; PRT; 299 AA.
Q8LKV3;
```

```
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Beta-carotene hydroxylase.
GN BCH1.
OS Vitis vinifera (Grape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Vitaceae; Vitis.
OX NCBI_TaxID=29760;
RN [1]_TaxID=29760;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Pinotage; TISSUE=Leaf;
RA Young P.R., Chen S.W., Vivier M.A.;
RT "Isolation, characterization and heterologous expression of a beta-
RT carotene hydroxylase from grapevine (Vitis vinifera).";
RE Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF493108; RAN77007.1; -.
DR GO: GO:0003824; P:catalytic activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005596; Carotene hydrox.
DR InterPro; IPR006087; Sterol desat.
DR Pfam; PF03897; Carotene hydrox. 1.
SQ SEQUENCE 299 AA; 32957 MW; 0231906399D21037 CRC64;

Query Match 27.5%; Score 263; DB 10; Length 299;
Best Local Similarity 41.2%; Pred. No. 1.8e-18;
Matches 54; Conservative 26; Mismatches 45; Indels 6; Gaps 3;

QY 15 VGMVEVAALAHKYMHGWMGWSHHEPRKGAFVNDLYAVVFAIVSIALIYFG--STG 72
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 145 VGMVEFWARWAKALWHSLSHHEPRKGAFVNDLYAVVFAIVSIALIYFG 204
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 73 IWPLOWIGAGM--TAYGLLYFMVHDGLVHQWPPRYIPRKGYLKRLYMAHMHAVRGKE 130
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 205 LVPLGCLFCGAGLGITVFGMAYNFVHDGLVHQWPPRYIPRKGYLKRLYMAHMHAVRGKE 262
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 131 GCVSFGFLYAP 141
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 263 NGVPYGLFLGP 273
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
Q8VXP2 PRELIMINARY; PRT; 305 AA.
AC Q8VXP2
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Beta-carotene hydroxylase.
GN BCH1CS.
OS Crocus sativus.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Iridaceae;
OC Crocus.
OX NCBI_TaxID=82528;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Stigma;
RA Gomez-Gomez L., Fernandez J.;
RT "Isolation and characterization of a carotenoid biosynthesis gene
RT coding for a carotenoid hydroxylase highly expressed in saffron
RT stigmas.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Stigma;
RA Gomez-Gomez L.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ416711; CAC95130.2; -.
SQ SEQUENCE 305 AA; 34108 MW; 42FFBE4BD45F870B CRC64;

Query Match 27.5%; Score 262.5; DB 10; Length 305;
```

```
Best Local Similarity 38.6%; Pred. No. 2.1e-18;
Matches 61; Conservative 29; Mismatches 55; Indels 13; Gaps 5;

Y 15 VGEVVAALAHKYMKGWGWHLSSHEPRKGFVNDLYAVVFAIVSIALIYFG--STG 72
b 146 VGEFEWARWAHRAHMAHSHWHSHSHRRPRGPFELNDVFAIINAVPALALNFGFFHRG 205
Y 73 IWPLOWIGAGM--TAYGLLYFMVHDCGLVHQRWPPRYIPRKGYLKRLYMAHRMHAHVROKE 130
b 206 LLEPLCFGAGLITLFGIAYMFVHDCGLVHRRFPVGPFIADVPYQFVAAAHOIHSEK-FE 264
Y 131 GCVSFGFLYAPP-----LSKQATLRERHARSAA 161
b 265 G-VPYGLFGPKLEBIEIGGLKELEKEVSRIRKAYNNSA 301

ESTULT 14
49815 PRELIMINARY; PRT; 315 AA.
D O49815
C O49815;
T 01-JUN-1998 (TRENBLrel. 06, Created)
T 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
T 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
E Beta-carotene hydrolase.
S Capsicum annuum (Bell pepper).
C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
C lamids; Solanales; Solanaceae; Capsicum.
X NCBI_TaxID=4072;
N [1]
N SEQUENCE FROM N.A.
X MEDLINE=98223452; PubMed=9555077;
JA Bouvier F., Keller Y., D'Harlingue A., Camara B.;
UT "Xanthophyll biosynthesis : molecular and functional characterization
UT of carotenoid hydroxylases from pepper fruits (Capsicum annuum L.).";
IL Biochim. Biophys. Acta 1391:320-328(1998).
JR EMBL; Y09225; CAA70427.1; -.
JR GO; GO:0016787; E:hydrolase activity; IEA.
JR GO; GO:0008152; P:metabolism; IEA.
JR InterPro; IPR005596; Carotene_hydrox.
JR InterPro; IPR006087; Sterol_desat.
JR Pfam; PF03897; Carotene_hydrox; 1.
CW Hydrolase.
SQ SEQUENCE 315 AA; 35426 MW; F68792194297E3B4 CRC64;

Query Match 27.3%; Score 261; DB 10; Length 315;
Best Local Similarity 35.4%; Pred. No. 3.1e-18;
Matches 54; Conservative 25; Mismatches 52; Indels 6; Gaps 3;

ZY 9 IVFVVVGEVVAALAHKYMKGWGWHLSSHEPRKGFVNDLYAVVFAIVSIALIYF 68
b 155 LAFGAAGMEVWARWAHRAHMAHSHWHSHSHRRPRGPFELNDVFAIINAVPALAFPSF 214
ZY 69 G--STGIWFLQWIGAGM--TAYGLLYFMVHDCGLVHQRWPPRYIPRKGYLKRLYMAHRMHH 124
b 215 GFNFKGLIPGICFGAGLITVFGMAFMVHDCGLVHRRFPVGPFIADVPYQFVAAAHOIHSEK 274
ZY 125 AVRGKGCVCVSFGFLYAP 141
b 275 S--DKFGVPGYGLFGP 289

RESULT 15
249814
ID O49814 PRELIMINARY; PRT; 316 AA.
AC O49814;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Beta-carotene hydroxylase 2 (fragment).
DS Capsicum annuum (Bell pepper).
DC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
DC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
```

```
OC lamids; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98223452; PubMed=9555077;
RA Bouvier F., Keller Y., D'Harlingue A., Camara B.;
RT "Xanthophyll biosynthesis : molecular and functional characterization
RT of carotenoid hydroxylases from pepper fruits (Capsicum annuum L.).";
RL Biochim. Biophys. Acta 1391:320-328(1998).
DR EMBL; Y09225; CAA70888.1; -.
DR GO; GO:0008152; P:metabolism; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR005596; Carotene_hydrox.
DR InterPro; IPR006087; Sterol_desat.
DR Pfam; PF03897; Carotene_hydrox; 1.
FT NON TER 1
SQ SEQUENCE 316 AA; 35792 MW; F59395B6DA552329 CRC64;

Query Match 27.2%; Score 260.5; DB 10; Length 316;
Best Local Similarity 38.0%; Pred. No. 3.5e-18;
Matches 57; Conservative 28; Mismatches 52; Indels 13; Gaps 4;

Oy 15 VGEVVAALAHKYMKGWGWHLSSHEPRKGFVNDLYAVVFAIVSIALIYFG--STG 72
Db 162 VGEFEWARWAHRAHMAHSHWHSHSHRRPRGPFELNDVFAIINAVPALALDYGFFHRG 221
Oy 73 IWPLOWIGAGM--TAYGLLYFMVHDCGLVHQRWPPRYIPRKGYLKRLYMAHRMHAHVROKE 130
Db 222 LIFGLCFGAGLITVFGMAFMVHDCGLVHRRFPVGPFIADVPYQFVAAAHOIHSEK 279
Oy 131 GCVSFGFLYAPP-----LSKQATLRER 153
Db 280 NGVPYGLFGPKLEBIEIGGLKELEKEVSRIR 309

Search completed: February 29, 2004, 14:51:08
Job time : 18.3897 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

on on: February 29, 2004, 14:27:18 ; Search time 2.94944 Seconds  
(without alignments)  
3089.496 Million cell updates/sec

tie: US-09-941-947a-36

fect score: 956

quence: 1 MLMWNALIVFTVVGMEVV.....ARSGAARDQGVDTSSSGK 175

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 141681 seqs, 52070155 residues

al number of hits satisfying chosen parameters: 141681

nimum DB seq length: 0

ximum DB seq length: 200000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	892	93.3	175	1 CRTZ_PANAN	P21688 pantoea ana
2	654	68.4	176	1 CRTZ_ERWHE	Q01332 erwinia her
3	459	52.2	162	1 CRTZ_ALCSP	Q42622 alcaligenes
4	495	51.8	162	1 CRTZ_AGRAU	P54973 agrobacteri
5	101.5	10.6	383	1 F03E_BRANA	P48624 brassica na
6	93.5	9.8	386	1 F03E_ARATH	P48623 arabidopsis
7	87	9.1	934	1 CAPP_MYCLE	P46710 mycobacteri
8	86.5	9.0	435	1 F03D_ARATH	P48622 arabidopsis
9	86	9.0	447	1 F03C_SESIN	P48620 sesamum ind
10	84.5	8.8	404	1 F03C_BRANA	P48618 brassica na
11	83	8.7	350	1 LSS1_HUMAN	P27544 homo sapien
12	82.5	8.6	446	1 F03C_ARATH	P46310 arabidopsis
13	82	8.6	662	1 CYOB_BUCAR	P57543 buchnera ap
14	81.5	8.5	460	1 F03C_RICCO	P48619 ricinus com
15	80	8.4	659	1 CYOB_BUCAR	Q8k994 buchnera ap
16	78	8.2	370	1 COMM_DROME	Q24139 drosophila
17	78	8.2	551	1 QCRB_MYCLE	P15878 mycobacteri
18	77.5	8.1	168	1 LSPA_VIBU	Q8des8 vibrio vuln
19	77.5	8.1	410	1 NUOH_MYCTU	P95174 mycobacteri
20	76.5	8.0	332	1 UL95_EBV	P03220 Epstein-Bar
21	76.5	8.0	453	1 F03C_SOYBN	P48621 glycine max
22	76.5	8.0	651	1 PPSA_PEDPE	P43470 pediococcus
23	76.5	8.0	712	1 NUOL_RHOCA	P50939 rhodobacter
24	76	7.9	542	1 YMB7_MYCTU	Q50678 mycobacteri
25	75.5	7.9	169	1 LSPA_VIBIO	Q87889 vibrio para
26	75.5	7.9	701	1 CSTA_ECOLI	P15078 escherichia
27	75	7.8	350	1 LSS1_MOUSE	P27545 mus musculu
28	75	7.8	1166	1 ADD8_BACSD	P23477 bacillus su
29	74	7.7	1032	1 Y450_SVNY3	P72637 synechocyst
30	72.5	7.6	431	1 UL78_HCMVA	P26751 human cytom
31	72	7.5	236	1 Y147_RICPR	Q9zel5 rickettsia
32	72	7.5	510	1 NANT_VERPE	Q8zch3 versinia pe
33	71.5	7.5	471	1 Y872_HAEIN	Q57491 haemophilus

34 71.5 7.5 527 1 MEMA\_METCA  
35 71.5 7.5 654 1 NUOL\_STROO  
36 71.5 7.5 946 1 Y2T6\_YEAST  
37 70.5 7.4 452 1 WZYE\_SALTI  
38 70.5 7.4 452 1 WZYE\_SALTY  
39 70.5 7.4 2436 1 ABC2\_HUMAN  
40 70 7.3 262 1 CYBH\_RHOCA  
41 69.5 7.3 171 1 LSPA\_VIBCH  
42 69.5 7.3 321 1 ALX\_ECOLI  
43 69.5 7.3 421 1 TRAB\_AGRTS  
44 69.5 7.3 424 1 PD6C\_SOYBN  
45 69.5 7.3 444 1 PSN\_CAEEL

## ALIGNMENTS

### RESULT 1

CRTZ\_PANAN STANDARD; PRT; 175 AA.

AC P21588:

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Beta-carotene hydroxylase.

GN CRTZ.

OS Pantoea ananas (Erwinia uredovora).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Pantoea.

OX NCBI\_TaxID=553;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=20D3;

RX MEDLINE=91072214; PubMed=2254247;

RA Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,

RA Nakamura K., Harashima K.;

RT "Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway

by functional analysis of gene products expressed in Escherichia

coli.";

RL J. Bacteriol. 172:6704-6712(1990).

CC -!- FUNCTION: Catalyzes the hydroxylation reaction from

beta-carotene to zeaxanthin.

CC -!- PATHWAY: Carotenoid biosynthesis.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D90087; BAA14129.1; -

DR PIR; F37802; F37802.

DR InterPro; IPR005596; Carotene\_hydrox.

DR InterPro; IPR006087; Sterol\_desat.

DR Pfam; PF03897; Carotene\_hydrox; 1.

DR Carotenoid biosynthesis.

SW SEQUENCE 175 AA; 19816 MW; 4BE1E5701CS1F9A CRC64;

Query Match 93.3%; Score 892; DB 1; Length 175;

Best Local Similarity 91.4%; Pred. No. 6,4e-79;

Matches 160; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 1 MLMWNALIVFTVVGMEVVAALAHKYIMHGCHGWHLSSHEPRKGAFFVNDLYAVVFAI 60

DB 1 MLMWNALIVFTVVGMEVVAALAHKYIMHGCHGWHLSSHEPRKGAFFVNDLYAVVFAA 60

QY 61 VSIALTYFGSTGIMPLWIGAGTAYGLLYFWHDGLVHQWPPFYIPRKGYLKLVMH 120

DB 61 LSILLIYLGSTGIMPLWIGAGTAYGLLYFWHDGLVHQWPPFYIPRKGYLKLVMH 120

QY 121 RMHVAVRGKGCVSFGFLYAPPLSKQLTLRERHAARSGAARDQGVDTSSSGK 175

[illegible]

```

DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Beta-carotene hydroxylase.
GN CRTZ.
OS Alcaligenes sp.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Alcaligenes.
OC NCBI_TaxID=512;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95251715; PubMed=7733978;
RA Misawa N., Kajiwara S., Kondo K., Yokoyama A., Satomi Y., Saito T.,
RA Miki W., Ohtani T.;
RT "Canthaxanthin biosynthesis by the conversion of methylene to keto
RL groups in a hydrocarbon beta-carotene by a single gene.";
RL Biochem. Biophys. Res. Commun. 209:867-876(1995).
CC -!- FUNCTION: Catalyzes the hydroxylation reaction from beta-carotene
CC to zeaxanthin via beta-cryptoxanthin (By similarity).
CC -!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin
CC biosynthetic pathway.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; D58422; BAA09597.1; ..
DR InterPro; IPR005596; Carotene_hydrox.
DR InterPro; IPR006087; Sterol_desat.
DR Pfam; PF03897; Carotene_hydrox; 1.
KW Carotenoid biosynthesis.
SQ SEQUENCE 162 AA; 19157 MW; 0CA5235CBEB33B5 CRC64;
-----
Query Match 52.2%; Score 499; DB 1; Length 162;
Best Local Similarity 57.8%; Pred. No. 4e-41;
Matches 93; Conservative 22; Mismatches 38; Indels 8; Gaps 4;

QY 8 LIVFVTVGKGVVAAAHKIKMHG-WGNGVHLSHHEPRKGAFVNDLYAVFAIVSIALI 66
   |||||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 5 LIVVATVIVNLTAYSVHRIMHGPFGWGNGHKSHEEHDALEKNDLYGVFVAVLATILP 64
   |||||:::||||:||||:||||:||||:||||:||||:||||:||||:

QY 67 YFGSTGIWP-LQWIGAGMTAYGLLYFVMDHGLVHQRWPFYIPRKGYLRLYVAHRMHA 125
   |||||:::||||:||||:||||:||||:||||:||||:||||:||||:
DB 65 TVGAYN-WPVLNWIALGTVTVGLIYFILLDELVHQRWPFYIIPRGYFRRLYQAHRLHA 123
   |||||:::||||:||||:||||:||||:||||:||||:||||:||||:

QY 126 VRGKEGCVGFGLYVAPLSKLQATLREHAAARGAARDEQ 166
   ||:::|||||:||||:||||:||||:||||:||||:||||:
DB 124 VEGRDHCVSPGFIYAPVVKLKQDLK-----RSGVLRPQDE 159
   ||:::|||||:||||:||||:||||:||||:||||:||||:

RESULT 4
CRTZ_AGRAU STANDARD; PRT; 162 AA.
ID CRTZ_AGRAU
AC P34973;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Beta-carotene hydroxylase.
GN CRTZ.
OS Agrobacterium aurantiacum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OC NCBI_TaxID=44155;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96062243; PubMed=7592436;
RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiwara S., Saito T.,
RA Ohtani T., Miki W.;
RT "Structure and functional analysis of a marine bacterial carotenoid
RT biosynthesis gene cluster and astaxanthin biosynthetic pathway

```

proposed at the gene level.";  
J. Bacteriol. 177:6575-6584(1995).  
-!- FUNCTION: Catalyzes the hydroxylation reaction from  
beta-carotene to zeaxanthin via beta-cryptoxanthin.  
-!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin  
biosynthetic pathway.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; D58420; BRA09592.1; -  
InterPro; IPR005596; Carotene hydros.  
InterPro; IPR006087; Sterol desat.  
Pfam; PF03897; Carotene hydros; 1.  
Carotenoid biosynthesis.  
SEQUENCE 162 AA; 18926 MW; F88891AA12A92C6C CRC64;  
-----  
Query Match 51.8%; Score 495; DB 1; Length 162;  
Best Local Similarity 58.0%; Pred. No. 9.7e-41;  
Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;  
-----  
6 NALLIVFTVWCVMEVVALAHKYIMHG-WKGNHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64  
3 NFLIVATVLMVELTAYSVHSHWIMHGLPMGKSHHEHDALEKNDLGLVFAIVATV 62  
-----  
65 LIYFGSTGIW- -LQWTCAGMTATGLLYFMVJGDLVHQRPFRIPKGYKLYMAHRM 122  
63 LFTVG- -WVAPVLMWALGNTVYGLIYFVLDGLVHQRPFRIPKGYARLYQARHL 120  
-----  
123 HVAVKGKGVGSFGLYAPPLSKLQATLRHRAARSGAARD 164  
121 HVAEGRDHCVSFGFIYAPPVVDKLDK-----MSGVLRAE 157  
-----  
SULT 5  
3E\_BRANA  
-----  
FD3E\_BRANA STANDARD; PRT; 383 AA.  
P48624; P46311;  
01-NOV-1995 (Rel. 32, Created)  
01-FEB-1996 (Rel. 33, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
Omega-3 fatty acid desaturase, endoplasmic reticulum (EC 1.14.19.-).  
FAD5.  
Brassica napus (Rapeseed).  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
eurosids II; Brassicales; Brassicaceae; Brassica.  
NCBI\_TaxID=3708;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=93088059; PubMed=1455229;  
Arondel V., Lemieux B., Hwang I., Gibson S., Goodman H.M.,  
Somerville C.R.;  
"Map-based cloning of a gene controlling omega-3 fatty acid  
desaturation in Arabidopsis.";  
Science 258:1353-1355(1992).  
[2]  
SEQUENCE FROM N.A.  
TISSUE=seed;  
MEDLINE=94302147; PubMed=8029334;  
Yadav N.S., Wierzbicki A., Aegerter M., Caster C.S., Perex-Grau L.,  
Kinney A.J., Hitz W.D., Booth J.R. Jr., Schweiger B., Stecca K.L.,  
Allen S.M., Blackwell M., Reiter R.S., Carlson T.J., Russell S.H.,  
Feldman K.A., Pierce J., Browne J.;  
"Cloning of higher plant omega-3 fatty acid desaturases.";  
Plant Physiol. 103:467-476(1993).  
-!- FUNCTION: ER (microsomal) omega-3 fatty acid desaturase introduces  
the third double bond in the biosynthesis of 18:3 fatty acids,

important constituents of plant membranes. It is thought to use  
cytochrome b5 as an electron donor and to act on fatty acids  
esterified to phosphatidylcholine and, possibly, other  
phospholipids.  
-!- PATHWAY: Polyunsaturated fatty acid biosynthesis.  
-!- SUBCELLULAR LOCATION: Endoplasmic reticulum.  
-!- DOMAIN: The histidine box domains may contain the active site  
and/or be involved in metal ion binding.  
-!- SIMILARITY: Belongs to the fatty acid desaturase family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; L01418; AAA32994.1; -  
EMBL; L22962; AAA61775.1; -  
PIR; A44227; A44227.  
InterPro; IPR005804; FA desat. fam.  
Pfam; PF00487; FA desaturase; 1.  
ProDom; PD001081; FA desat. fam; 2.  
Oxidoreductase; Fatty acid biosynthesis; Endoplasmic reticulum;  
Transmembrane.  
PT TRANSMEM 53 73 POTENTIAL.  
FT TRANSMEM 210 230 POTENTIAL.  
FT TRANSMEM 234 254 POTENTIAL.  
FT DOMAIN 98 102 HISTIDINE BOX-1.  
FT DOMAIN 134 138 HISTIDINE BOX-2.  
FT DOMAIN 301 305 HISTIDINE BOX-3.  
FT CONFLICT 11 11 V -> A (IN REF. 2).  
FT CONFLICT 15 22 SGARKEBG -> ER (IN REF. 2).  
FT CONFLICT 57 57 T -> A (IN REF. 2).  
FT CONFLICT 64 64 A -> V (IN REF. 2).  
FT CONFLICT 68 68 M -> V (IN REF. 2).  
FT CONFLICT 78 78 L -> F (IN REF. 2).  
FT CONFLICT 84 84 V -> A (IN REF. 2).  
FT CONFLICT 113 114 SV -> TA (IN REF. 2).  
FT CONFLICT 162 162 P -> S (IN REF. 2).  
FT CONFLICT 181 181 I -> L (IN REF. 2).  
FT CONFLICT 195 195 P -> Y (IN REF. 2).  
FT CONFLICT 232 232 D -> G (IN REF. 2).  
FT CONFLICT 264 264 E -> D (IN REF. 2).  
FT CONFLICT 320 321 RA -> KS (IN REF. 2).  
SQ SEQUENCE 383 AA; 43936 MW; DAD7C3AG67A12826A CRC64;  
-----  
Query Match 10.6%; Score 101.5; DB 1; Length 383;  
Best Local Similarity 25.5%; Pred. No. 0.018;  
Matches 26; Conservative 20; Mismatches 29; Indels 27; Gaps 5;  
-----  
QY 44 RKGAPEVNDLYAVVFAIVSIALIYFGSTGIWPLQWICAGMTATGLLYFMVHD----- 95  
DB 51 RMSYVTRDIPAV--ALAAVAIVFDSWFLNPLYWVAQG-TLFWAIFVLGDCGHGSFSD 107  
-----  
QY 96 -----GLVHQRPFRIPKGYKLYMAHRMHAVRG 128  
DB 108 IPLLNSVVGHILHS---FILVPYHGW----RISHRTTHQNHG 142  
-----  
RESULT 6  
FD3E\_ARATH  
ID FD3E\_ARATH STANDARD; PRT; 386 AA.  
AC P48623;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Omega-3 fatty acid desaturase, endoplasmic reticulum (EC 1.14.19.-).  
GN FAD3 OR AT2G29980 OR F23F1.10.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;



Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
Rutter S., Seeger K., Simon S., Simmonds M., Sketton J., Squares R.,  
Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
Barrell B.G.;  
"Massive gene decay in the leprosy bacillus";  
Nature 409:1007-1011(2001).  
-!- FUNCTION: Through the carboxylation of phosphoenolpyruvate (PEP)  
it forms oxaloacetate, a four-carbon dicarboxylic acid source for  
the tricarboxylic acid cycle.  
-!- CATALYTIC ACTIVITY: Phosphate + oxaloacetate = H(2)O +  
phosphoenolpyruvate + CO(2).  
-!- PATHWAY: Tricarboxylic acid cycle.  
-!- SIMILARITY: Belongs to the PEPCase family.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).

-----

EMBL; U00013; AAA17132.1; ALT INIT.  
EMBL; AL583919; CAC30086.1; -  
PIR; B86981; B86981.  
HSP; P00864; 1FTY.  
Leptoma; MLQ578; -  
HAMAP; MF\_00595; -; 1.  
InterPro; IPRO01449; PEPcase.  
Pfam; PF00311; PEPcase; 1.  
PRINTS; PR00150; PEPCARXLAZE.  
PROSITE; PS00393; PEPCASE 2; 1.  
PROSITE; PS00781; PEPCASE 1; 1.  
Lyase; Carbon dioxide fixation; Tricarboxylic acid cycle;  
Complete proteome.  
ACT\_SITE 161 161 BY SIMILARITY.  
ACT\_SITE 593 593 BY SIMILARITY.  
SEQUENCE 934 AA; 102515 MW; 3EE8FD762EC45180 CRC64;

Query Match 9.1%; Score 87; DB 1; Length 934;  
Best Local Similarity 23.5%; Pred No. 1.1;  
Matches 38; Conservative 21; Mismatches 53; Indels 50; Gaps 7;

44 RKGAPEVNDLYAVVFAIV---SIATLYFGSTGIPLOWIGMGMTAYGLLYFWVDGLVH 99  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
752 RXPPTSIAIDLRAIPWLAWSQRWMLPGWYGTSAPQQWAAGPSQSQRVELMD--LY 809  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
100 QWPFP-----RYTP-----RKYLKRLYYAHR-----MEHAVER 127  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
810 QWPFPFRVLNMAQVLAKSDGLAARYAEVLVDEALRRRVDFDIADREHTTAHKLI 669  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
128 KGEGCVSFQ-----FLYAPPLSKQLATREFHAARSQ 159  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :  
870 GHDDLADNPALARSFVRFPYLELNHLQVELLRKY--RSQ 909  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| :

SULT 8  
D ADARATH STANDARD; PRT; 435 AA.  
P48622;  
01-FEB-1996 (Rel. 33, Created)  
01-FEB-1996 (Rel. 33, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
Temperature-sensitive omega-3 fatty acid desaturase, chloroplast  
precursor (EC 1.14.19.-).  
FAD8 OR AT5G05580 OR MOP10.12.  
Arabidopsis thaliana (Mouse-ear cress).  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
eucoids 11; Brassicales; Brassicaceae; Arabidopsids.  
NCBI\_TaxId=3702;  
[1]  
SEQUENCE FROM N.A.

```

RC STRAIN=cv. Columbia; TISSUE=Aerial parts;
RX MEDLINE=95148742; PubMed=7846164;
RG Gibson S., Arondel V., Iba K., Somerville C.R.;
RT "Cloning of a temperature-regulated gene encoding a chloroplast
RT omega-3 desaturase from Arabidopsis thaliana.";
RL Plant Physiol. 106:1615-1621(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Columbia; TISSUE=Hypocotyl;
RC Watahiki M.C., Yamamoto K.T.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RC MEDLINE=97471969; PubMed=9330910;
RG Sato S., Kotani H., Nakamura Y., Kaneko T., Asanizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT pi clones.";
RL DNA Res. 4:215-230(1997).
RC -1- FUNCTION: Chloroplast omega-3 fatty acid desaturase introduces the
CC third double bond in the biosynthesis of 16:3 and 18:3 fatty
CC acids, important constituents of plant membranes. It is thought to
CC use ferredoxin as an electron donor and to act on fatty acids
CC esterified to galactolipids, sulfolipids and phosphatidylglycerol.
CC -1- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).
CC -1- INDUCTION: BY LOW TEMPERATURE.
CC -1- DOMAIN: The histidine box domains may contain the active site
CC and/or be involved in metal ion binding.
CC -1- SIMILARITY: Belongs to the fatty acid desaturase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L27158; AAA5621.1; --
CC DR EMBL; U08216; AAB50302.1; --
CC DR EMBL; D17578; BAA04504.1; --
CC DR EMBL; AB005241; BAB11547.1; --
CC DR InterPro; IPR005804; FA_desat fam.
CC DR Pfam; PF00487; FA_desaturase_1.
CC DR ProDom; PD001081; FA_desat fam; 2.
CC DR Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW Transit peptide.
KW TRANSIT ?
KW CHAIN ? 435 ? CHLOROPLAST (POTENTIAL).
FT TEMPERATURE-SENSITIVE OMEGA-3 FATTY ACID
FT DESATURASE.
FT DOMAIN 156 160 HISTIDINE BOX-1.
FT DOMAIN 192 196 HISTIDINE BOX-2.
FT DOMAIN 359 363 HISTIDINE BOX-3.
FT SEQUENCE 435 AA; 50136 MW; 3D77A8035A6214E1 CRC64;
SQ
Query Match 9.0%; Score 86.5; DB 1; Length 435;
Best Local Similarity 26.3%; Pred. No. 0.56;
Matches 26; Conservative 17; Mismatches 29; Indels 27; Gaps 6;
QY 47 AFEVDLYAVFAIVSIALYLYFGSGTGIWPLQWIGAGMTAYGLLYFMVHD----- 95
DB 112 SYVREDV-AIVFGLAAVA-AYFNWLLNPLYPFAQG-TMFWALFVLGHDCGHGSFSDNDR 168
QY 96 -----GLVHORWPPRYIPRKGYLKRLVMAHMHMAVRG 128
DB 169 LNSVAGHLHSS---LLVPEYHGW----RISHRTHEQNEG 200

```

D FD3C SESIN STANDARD; PRT; 447 AA.  
 C P48620;  
 T 01-FEB-1996 (Rel. 33, Created)  
 T 01-FEB-1996 (Rel. 33, Last sequence update)  
 T 28-FEB-2003 (Rel. 41, Last annotation update)  
 E Omega-3 fatty acid desaturase, chloroplast precursor (EC 1.14.19.-).  
 N NCBI\_TaxID=4182;  
 S Sesamum indicum (Oriental sesame) (Gingelly).  
 C Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 C Lamiales; Pedaliaceae; Sesamum.  
 X NCBI\_TaxID=4182;  
 [1]  
 N SEQUENCE FROM N.A.  
 P STRAIN=CV. 4294; TISSUE=Cotyledon;  
 A Shoji K.;  
 L Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 C -1- FUNCTION: Chloroplast omega-3 fatty acid desaturase introduces the  
 C third double bond in the biosynthesis of 16:3 and 18:3 fatty  
 C acids, important constituents of plant membranes. It is thought to  
 C use ferredoxin as an electron donor and to act on fatty acids  
 C esterified to galactolipids, sulfolipids and phosphatidylglycerol.  
 C -1- PATHWAY: Polyunsaturated fatty acid biosynthesis.  
 C -1- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).  
 C -1- DOMAIN: The histidine box domains may contain the active site  
 C and/or be involved in metal ion binding.  
 C -1- SIMILARITY: Belongs to the fatty acid desaturase family.  
 C  
 C This SWISS-PROT entry is copyright. It is produced through a collaboration  
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 C the European Bioinformatics Institute. There are no restrictions on its  
 C use by non-profit institutions as long as its content is in no way  
 C modified and this statement is not removed. Usage by and for commercial  
 C entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 C  
 C EMBL: U25817; AAA70334.1; -;  
 C InterPro: IPR005804; FA\_desat\_fam.  
 C Pfam: PF00487; FA\_desaturase\_1.  
 C ProDom: PD001081; FA\_desat\_fam; 2.  
 C Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;  
 C Transit peptide.  
 C TRANSIT 1 ? CHLOROPLAST (POTENTIAL).  
 C CHAIN 1 447 OMEGA-3 FATTY ACID DESATURASE.  
 C DOMAIN 167 171 HISTIDINE BOX-1.  
 C DOMAIN 203 207 HISTIDINE BOX-2.  
 C DOMAIN 370 374 HISTIDINE BOX-3.  
 C SEQUENCE 447 AA; 51116 MW; 4E76250DD6DA6B1 CRC64;  
 C  
 C Query Match 9.0%; Score 86; DB 1; Length 447;  
 C Best Local Similarity 24.0%; Pred. No. 0.64;  
 C Matches 29; Conservative 18; Mismatches 38; Indels 38; Gaps 7;  
 Y 25 HKYIMHGCGWGLSHHSHPRKGFVNDLYAVVFAIVSTALYFGSTGIWPLWIGAGMT 84  
 b 112 HCWVKDPW-----RSMGVYVRDY-AVYFGLAAVA-AYFNWVWVWPLWYF-AQST 157  
 Y 85 AYGLLPMVHD-----GLVHQRPFRYPYPRKGLKELLYNARHGHAVR 127  
 b 158 MFVALFVLGHDCGHSFNDPKLNSVVGHLHSS-----ILVPYHGW----RISHRTHQNH 210  
 Y 128 G 128  
 b 211 G 211  
 RESULT 10  
 D3C BRANA STANDARD; PRT; 404 AA.  
 C P48618;  
 T 01-FEB-1996 (Rel. 33, Created)  
 T 01-FEB-1996 (Rel. 33, Last sequence update)  
 T 28-FEB-2003 (Rel. 41, Last annotation update)

DE Omega-3 fatty acid desaturase, chloroplast precursor (EC 1.14.19.-)  
 DN {Fragment}.  
 GN FAD7  
 OS Brassica napus (Rape).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 CC eurosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN NCBI\_TaxID=3708;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Seed.  
 RX MEDLINE=94302147; PubMed=8029334;  
 RA Yadav N.S., Wierzbicki A., Aegerter M., Caster C.S., Perez-Grau L.,  
 RA Kinney A.J., Hitz W.D., Booth J.R. Jr., Schweiger B., Stecca K.L.,  
 RA Allen S.M., Blackwell M., Reiter R.S., Carlson T.J., Russell S.H.,  
 RA Feldmann K.A., Pierce J., Browse J.;  
 RT "Cloning of higher plant omega-3 fatty acid desaturases";  
 RL Plant Physiol. 103:467-476(1993).  
 CC -1- FUNCTION: Chloroplast omega-3 fatty acid desaturase introduces the  
 CC third double bond in the biosynthesis of 16:3 and 18:3 fatty  
 CC acids, important constituents of plant membranes. It is thought to  
 CC use ferredoxin as an electron donor and to act on fatty acids  
 CC esterified to galactolipids, sulfolipids and phosphatidylglycerol.  
 CC -1- PATHWAY: Polyunsaturated fatty acid biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).  
 CC -1- DOMAIN: The histidine box domains may contain the active site  
 CC and/or be involved in metal ion binding.  
 CC -1- SIMILARITY: Belongs to the fatty acid desaturase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).)  
 CC  
 CC EMBL: L22963; AAA61774.1; ALT\_INIT.  
 CC PIR: PQ0812; PQ0812.  
 CC InterPro: IPR005804; FA\_desat\_fam.  
 CC Pfam: PF00487; FA\_desaturase\_1.  
 CC ProDom: PD001081; FA\_desat\_fam; 2.  
 CC Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;  
 CC Transit peptide.  
 CC NON TER 1 1 CHLOROPLAST (POTENTIAL).  
 C TRANSIT <1 404 OMEGA-3 FATTY ACID DESATURASE.  
 C CHAIN ? 404 HISTIDINE BOX-1.  
 C DOMAIN 121 125 HISTIDINE BOX-2.  
 C DOMAIN 157 161 HISTIDINE BOX-3.  
 C DOMAIN 324 328 HISTIDINE BOX-3.  
 C SEQUENCE 404 AA; 46617 MW; 4B59FB2F36E2EDE4 CRC64;  
 C  
 C Query Match 8.8%; Score 84.5; DB 1; Length 404;  
 C Best Local Similarity 24.0%; Pred. No. 0.8;  
 C Matches 31; Conservative 23; Mismatches 36; Indels 39; Gaps 8;  
 Y 18 EVVAAL-AHKYIMHGCGWGLSHHSHPRKGFVNDLYAVVFAIVSTALYFGSTGIWPL 76  
 b 58 DIRAALPKCKGKQKPN-----KSMGVYVRDY-AVYFGLAAVA-AYFNWVWVWPL 104  
 Y 77 QWIGAGTAYGLLYPMVHD-----GLVHQRPFRYPYPRKGLKELLYNAR 119  
 b 105 YWTAQG-TMFWALFVLGHDCGHSFNDPKLNSVVGHLHSS-----ILVPYHGW----RIS 156  
 Y 120 HRMHAVRG 128  
 b 157 HRTHQNHG 165  
 RESULT 11  
 LSSI HUMAN  
 ID LSSI HUMAN STANDARD; PRT; 350 AA.  
 AC P27544;

01-AUG-1992 (Rel. 23, Created)  
01-AUG-1992 (Rel. 23, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
Longevity assurance homolog 1 (UOG-1 protein) (LAG1 protein).  
LASSI OR UOG1 OR LAG1.  
Homo sapiens (Human).  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A. (ISOFORM 1).  
MEDLINE=91239545; PubMed=2034669;  
Lee S.-J.;  
"Expression of growth/differentiation factor 1 in the nervous system:  
conservation of a bicistronic structure."  
Proc. Natl. Acad. Sci. U.S.A. 88:4250-4254(1991).  
[2]  
SEQUENCE FROM N.A. (ISOFORM 1).  
MEDLINE=93091901; PubMed=9872981;  
Jiang J.C., Kirchner P.A., Zeglinski M., Hunt J., Jaswinski S.M.;  
"Homologs of the yeast longevity gene LAG1 in Caenorhabditis elegans  
and human."  
Genome Res. 8:1259-1272(1998).  
[3]  
SEQUENCE FROM N.A.  
Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,  
Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,  
Phan H., Velasco N., Do L., Regala W., Terry A., Ganes J.,  
Danganan L., Poundstone P., Christensen M., Georgescu A., Avila J.,  
Liu S., Attix C., Andreise T., Frankheim M., Amico-Keller G.,  
Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G.,  
Kronmiller B., Arellano A., Montgomerly M., Ow D., Nolan M., Trong S.,  
Kobayashi A., Olsen A.S., Carrano A.V.;  
Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE FROM N.A. (ISOFORM 2).  
TISSUE=Hypothalamus;  
MEDLINE=22389257; PubMed=12477932;  
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalius D.E.,  
Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences."  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
-!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
-!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Name=1;  
IsoId=927544-1; Sequence=Displayed;  
Name=2;  
IsoId=927544-2; Sequence=VSP\_003049;  
Note=No experimental confirmation available;  
-!- SIMILARITY: Contains 1 T1C (TRAM/LAG1/CLN8) domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
DR EMBL; M62302; AAA58500.1; --  
DR EMBL; AF105009; AAD16892.1; --  
DR EMBL; AF105005; AAD16892.1; JOINED.  
DR EMBL; AF105006; AAD16892.1; JOINED.  
DR EMBL; AF105007; AAD16892.1; JOINED.  
DR EMBL; AF105008; AAD16892.1; JOINED.  
DR EMBL; AF105009; AAD16892.1; JOINED.  
DR EMBL; AC005197; AAC24611.1; --  
DR EMBL; AC003972; -- NOT ANNOTATED\_CDS.  
DR EMBL; BC022450; AAH22450.1; --  
DR PIR; D39364; D39364.  
DR Genew; HGNC:14253; LASS1.  
DR MIM; 606919; --  
DR GO; GO:0016021; C: integral to membrane; ISS.  
DR GO; GO:0007568; P: aging; ISS.  
DR InterPro; IPR005547; LAG1.  
DR InterPro; IPR006634; T1C.  
DR Pfam; PF03798; LAG1; 1.  
DR SMART; SM00734; T1C; 1.  
DR PROSITE; PS00922; T1C; 1.  
KW Transmembrane; Alternative splicing.  
FT TRANSMEM 53 73 POTENTIAL.  
FT TRANSMEM 103 123 POTENTIAL.  
FT TRANSMEM 148 168 POTENTIAL.  
FT TRANSMEM 176 196 POTENTIAL.  
FT TRANSMEM 239 259 POTENTIAL.  
FT TRANSMEM 287 307 POTENTIAL.  
FT DOMAIN 97 311 T1C.  
FT VARSPLIC 338 350 Missing (in isoform 2).  
FT COMPLET 111 111 /FTID=VSP\_003049.  
FT SEQUENCE 350 AA; 39536 MW; F102C12C47DB4162 CRC64;  
SQ  
Query Match 8.7%; Score 83; DB 1; Length 350;  
Best Local Similarity 22.2%; Pred. No. 0.97;  
Matches 42; Conservative 18; Mismatches 53; Indels 76; Gaps 8;  
QY 3 WINNALIVFVT-----VVGMEVVAALAHKYMIGWGWG-----W 36  
DB 113 WSAYALLFGDYDFHDPFVFDVTPGMAVPRDIAAYLLQGSFYGHSHIYATLYMDTW 172  
QY 37 H-----LSHHEPRKGAFEVNDLVAVVFAIVSIALIYFGSTGIWPLQMGAGMTAYGLLY 90  
DB 173 REDSVMLLHH-----VVTLLIVSSVAFRYEN-----VGILV 205  
QY 91 FMVHD-----GLVHORMFPRIPKGYLKRLYMHRMHEAVRGEGCVSPF-----LYA 140  
DB 206 LFLHDSIDVQLFETKLNIYFKSRGGSYHRL-----HALAADIGCLSGFSGFWFELYW 258  
QY 141 PPLSKLOAT 149  
DB 259 PPLKVLVAT 267  
RESULT 12  
PD3C ARATH  
ID PD3C ARATH STANDARD; PRT; 446 AA.  
AC P46310;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Omega-3 fatty acid desaturase, chloroplast precursor (EC 1.14.19.-).  
OS FAD7 OR FADD OR AT3G1170 OR F9F8.4 OR F11B9.10.  
QS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia; TISSUE=Hypocotyl;

MEDLINE=94302147; PubMed=8029334;  
 Yadav N.S., Wierzbicki A., Aegerter M., Caster C.S., Perez-Grau L.,  
 Kinney A.J., Hitz W.D., Booth R.R. Jr., Schweiger B., Stecca K.L.,  
 Allen S.M., Blackwell M., Reiter R.J., Carlson T.J., Russell S.H.,  
 Faldmann K.A., Pierce J., Browne J.,  
 "Cloning of higher plant omega-3 fatty acid desaturases.",  
 Plant Physiol. 103:467-476(1993).  
 [2]  
 P SEQUENCE FROM N.A.  
 C STRAIN=cv. Columbia; TISSUE=Aerial parts;  
 MEDLINE=94043239; PubMed=8226956;  
 Iba K., Gibson S., Nishitani T., Fuse T., Nishimura M., Arondel V.,  
 Hugly S., Somerville C.R.;  
 "A gene encoding a chloroplast omega-3 fatty acid desaturase  
 complements alterations in fatty acid desaturation and chloroplast  
 copy number of the fad7 mutant of Arabidopsis thaliana.",  
 J. Biol. Chem. 268:24099-24105(1993).  
 [3]  
 N SEQUENCE FROM N.A.  
 C STRAIN=cv. Columbia; TISSUE=Hypocotyl;  
 Watabiki M., Yamamoto K.;  
 Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 [4]  
 N SEQUENCE FROM N.A.  
 C STRAIN=cv. Columbia;  
 MEDLINE=21016720; PubMed=11130713;  
 A Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unselid M.,  
 Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 Dalseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,  
 De Simone V., Choise N., Attiguenave F., Robert C., Brottier P.,  
 Wincker P., Cattolico L., Weissensbach J., Saurin W., Quetier F.,  
 Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 Wurmbach E., Drzonek H., Erfle H., Holland R., Brandt S.,  
 Wiedemann R., Kranz H., Voss H., Jorand N., Bangerter S.,  
 Verzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,  
 Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,  
 Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,  
 Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,  
 Cooke R., Laudie M., Berger-Liauro C., Purnelle B., Masuy D.,  
 De Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,  
 Monfort A., Argireou A., Flores M., Liguori R., Vitale D.,  
 Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,  
 Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,  
 Creasy T.H., Haas B., Maiti A., Wu D., Peterson J., Van Aken S.,  
 Pal G., Millscher J., Sellers P., Gill J.E., Feldblyum T.V.,  
 Pruss D., Lin X., Niemman W.C., Salzberg S.L., White O., Venter J.C.,  
 Fraser C.M., Kaneo T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
 Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
 Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Muraki A.,  
 Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,  
 Watanabe A., Yamada M., Yasuda M., Tabata S.;  
 "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 thaliana.",  
 Nature 408:820-822(2000).  
 IL -!- FUNCTION: Chloroplast omega-3 fatty acid desaturase introduces  
 the third double bond in the biosynthesis of 16:3 and 18:3 fatty  
 acids, important constituents of plant membranes. It is thought  
 to use ferredoxin as an electron donor and to act on fatty acids  
 esterified to galactolipids, sulfolipids and phosphatidylglycerol.  
 -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.  
 -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).  
 -!- TISSUE SPECIFICITY: Most abundant in leaves and seedlings.  
 -!- DOMAIN: The histidine box domains may contain the active site  
 and/or be involved in metal ion binding.  
 -!- SIMILARITY: Belongs to the fatty acid desaturase family.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC EMBL; L22961; AAR61773.1; -  
 DR EMBL; D14007; BAA03106.1; -  
 DR EMBL; D26019; BAA05040.1; -  
 DR EMBL; AC009991; AAF01508.1; -  
 DR EMBL; AC073395; AAG50977.1; -  
 DR PIR; JQ2336; JQ2336.  
 DR InterPro: IPR005804; FA desat fam.  
 DR Pfam: PF00487; FA desaturase; 1.  
 DR ProDom: PD001081; FA desat fam; 2.  
 KM Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;  
 KM Transit peptide.  
 FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).  
 FT CHAIN 1 446 OMEGA-3 FATTY ACID DESATURASE.  
 FT DOMAIN 163 167 HISTIDINE BOX-1.  
 FT DOMAIN 199 203 HISTIDINE BOX-2.  
 FT DOMAIN 366 370 HISTIDINE BOX-3.  
 SQ SEQUENCE 446 AA; 51174 MW; 121125F634553D35 CRC64;  
 Query Match 8.6%; Score 82.5; DB 1; Length 446;  
 Best Local Similarity 23.3%; Pred. No. 1.4;  
 Matches 30; Conservative 24; Mismatches 36; Indels 39; Gaps 8;  
 Qy 18 EVVAAL-AHKYIMGKMGWHLSSHEPRKGAPEVNDLVAVFAIVSIALIVFGSTGIMPL 76  
 Db 100 DIFAAIPKHCWKNPW-----KLSYVVRDV-AIVFALAAGA-AYLNWIVWPL 146  
 Qy 77 QMTGAGTAYGLLYFMVHD-----GLVHQRPFRYPYPRKGYLRLYMA 119  
 Db 147 YMLAQG-TMFVALFVLGHGCHGHSFNSDKLSNVVGHLLHSS---ILVPHGW----RIS 198  
 Qy 120 HRVHVAVRG 128  
 Db 199 HRTTHQMHG 207  
 RESULT 13  
 CYOB\_BUCAI STANDARD; PRT; 562 AA.  
 ID CYOB\_BUCAI  
 AC P57543;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ubiquinol oxidase polypeptide I (EC 1.10.3.-) (Cytochrome O subunit 1)  
 DE (Oxidase BO(3) subunit 1) (Cytochrome O ubiquinol oxidase subunit 1).  
 DE CYOB OR B0471.  
 GN Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 symbiotic bacterium).  
 OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 OX NCBI\_TaxID=118099;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998;  
 RA MEDLINE=20445173; PubMed=10993077;  
 RX Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "Genome sequence of the endocellular bacterial symbiont of aphids  
 Buchnera sp. APS.",  
 RL Nature 407:81-85(2000).  
 CC -!- FUNCTION: Cytochrome O terminal oxidase complex is the component  
 of the aerobic respiratory chain that predominates when cells are  
 grown at high aeration. This ubiquinol oxidase shows proton pump  
 activity across the membrane in addition to the electron transfer  
 (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Ubiquinol-8 + O(2) = Ubiquinone-8 + H(2)O.  
 CC -!- COFACTOR: Contains two protoheme IX (heme B55 and B562) and copper  
 B (By similarity).  
 CC -!- PATHWAY: Ubiquinol oxidase catalyzes the terminal step in the  
 electron transport chain.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -!- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.



This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
EMBL; AP001119; BAB13168.1; -
HSSP; P18401; 1PFT.
InterPro; IPR000883; COX1.
Pfam; PF00115; COX1.1.
PRINTS; PR01165; CYCOXIDASSI.
PROSITE; PS00077; COX1.1.
Oxidoreductase; Respiratory chain; Heme; Copper;
Hydrogen ion transport; Complete;
DOMAIN 1 14
TRANSMEM 15 35
EXTRACELLULAR (POTENTIAL).
DOMAIN 36 58
CYTOPLASMIC (POTENTIAL).
DOMAIN 59 79
TRANSMEM 80 106
EXTRACELLULAR (POTENTIAL).
DOMAIN 107 127
TRANSMEM 128 145
CYTOPLASMIC (POTENTIAL).
DOMAIN 146 166
EXTRACELLULAR (POTENTIAL).
DOMAIN 167 189
TRANSMEM 190 210
EXTRACELLULAR (POTENTIAL).
DOMAIN 211 232
CYTOPLASMIC (POTENTIAL).
DOMAIN 233 253
TRANSMEM 254 277
EXTRACELLULAR (POTENTIAL).
DOMAIN 278 298
TRANSMEM 299 309
CYTOPLASMIC (POTENTIAL).
DOMAIN 310 330
TRANSMEM 331 346
EXTRACELLULAR (POTENTIAL).
DOMAIN 347 367
TRANSMEM 368 380
CYTOPLASMIC (POTENTIAL).
DOMAIN 381 401
TRANSMEM 402 413
EXTRACELLULAR (POTENTIAL).
DOMAIN 414 434
TRANSMEM 435 456
CYTOPLASMIC (POTENTIAL).
DOMAIN 457 477
TRANSMEM 478 493
EXTRACELLULAR (POTENTIAL).
DOMAIN 494 514
TRANSMEM 515 586
CYTOPLASMIC (POTENTIAL).
DOMAIN 587 607
TRANSMEM 608 608
EXTRACELLULAR (POTENTIAL).
DOMAIN 609 629
TRANSMEM 630 662
CYTOPLASMIC (POTENTIAL).
DOMAIN 106 106
IRON (HEME B AXIAL LIGAND) (PROBABLE).
METAL 284 284
COPPER B (PROBABLE).
METAL 288 288
COPPER B (PROBABLE).
METAL 333 333
COPPER B (PROBABLE).
METAL 334 334
COPPER B (PROBABLE).
METAL 419 419
IRON (HEME O AXIAL LIGAND) (PROBABLE).
METAL 421 421
IRON (HEME B AXIAL LIGAND) (PROBABLE).
1'-histidyl-3'-tyrosine (His-Tyr)
(BY similarity).
CROSSLNK 284 288
SEQUENCE 662 AA; 75455 MW; 3F502A628133EA65 CRC64;
Query Match 8.6%; Score 82; DB 1; Length 662;
Best Local Similarity 20.4%; Pred. No. 2.3;
Matches 39; Conservative 25; Mismatches 53; Indels 74; Gaps 12;
1 MLTWNALIVFTV-----VGMVEVAALAHK-----YTMHONG-----WGNHLSH 40
278 LIIWGHPEVILVLPFGVPSEVATFSKRLPGYSLVNATILSIFVW-----LH 333
41 HEPKGA-FEVNDIYAVVPAIVSIALIVFGSTGWPLOWIGAGTAYGLLYFNVHDLVH 99
334 HFTVAGADVNTFFGTTMLIAL-----PTGVKIFNWL-----FTIYQGRVH 376
100 QR-----WPFY-----IPKGYL--KRLYNAMRHMAVRGKE--GCVS-- 134
377 MESSILWTLGFLVTFISIGMTGVLVSPADPFLVHNSLFLVAHFHNVILGGVVFGC FAGI 436
```

```
QY 135 -----EGFL 138
Db 437 NYWFKLFGFV 447
RESULT 14
FD3C RICCO STANDARD; PRT; 460 AA.
AC P48619;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Omega-3 fatty acid desaturase, chloroplast precursor (EC 1.14.19.-).
GN PAD7A-1.
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosids;
OC Eurosids I; Malpighiales; Euphorbiaceae; Acalyphoideae; Acalyphaceae;
OC Ricinus.
OX NCBI_TaxID=3998;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=cv. Baker 296; TISSUE=Seed;
MEDLINE=94302177; PubMed=8029360;
van de Ioo F.J.; Somerville C.R.;
"Plasmid omega-3 fatty acid desaturase cDNA from Ricinus communis.";
Plant Physiol. 195; 443-444 (1994).
CC -!- FUNCTION: Chloroplast omega-3 fatty acid desaturase introduces the
third double bond in the biosynthesis of 16:3 and 18:3 fatty
acids, important constituents of plant membranes. It is thought to
use ferredoxin as an electron donor and to act on fatty acids
esterified to galactolipids, sulfolipids and phosphatidylglycerol.
CC -!- PATHWAY: Polyunsaturated fatty acid biosynthesis.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST, MEMBRANE-BOUND (PROBABLE).
CC -!- DOMAIN: The histidine box domains may contain the active site
and/or be involved in metal ion binding.
CC -!- SIMILARITY: Belongs to the fatty acid desaturase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
EMBL; L25897; AAA73511.1; -
PIR; T10063; T10063.
InterPro; IPR005804; FA desat fam.
Pfam; PF00487; FA desaturase; 1.
ProDom; PD001081; FA desat fam; 2.
KW Oxidoreductase; Fatty acid biosynthesis; Chloroplast; Membrane;
KW Transit peptide.
FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
FT CHAIN ? 460 OMEGA-3 FATTY ACID DESATURASE.
FT DOMAIN 177 181 HISTIDINE BOX-1.
FT DOMAIN 213 217 HISTIDINE BOX-2.
FT DOMAIN 380 384 HISTIDINE BOX-3.
SQ SEQUENCE 460 AA; 52561 MW; 836592904E93C7B0 CRC64;
Query Match 8.5%; Score 81.5; DB 1; Length 460;
Best Local Similarity 23.3%; Pred. No. 1.8;
Matches 30; Conservative 22; Mismatches 38; Indels 39; Gaps 8;
QY 18 EVVAL-AHKYIMHGKMGWELSHHEPRKGAFFVNDIYAVVPAIVSIALIVFGSTGWPL 76
Db 114 DIRAIPKHCWKPW-----RSMYSVLRDV-VVFGLAAVA-AYFNWVAVPL 160
QY 77 QWIGAGTAYGLLYFNVHD-----GLVHORWPFYIPKGYLKRLYMA 119
Db 161 YWFCQ-TMFWALFVLGHDCHGHSFNNPKLNSVVGHLHSS---ILVPYHGW---RIS 212
```

```

2Y 120 HRMHAVRG 128
2b 213 HRTHQNHG 221

RESULT 15
YOB_BUCAP
D YOB_BUCAP STANDARD; PRT; 659 AA.
AC 08K994;
KT 28-FEB-2003 (Rel. 41, Created)
YR 28-FEB-2003 (Rel. 41, Last sequence update)
YR 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ubiquinol oxidase polypeptide I (EC 1.10.3.-) (Cytochrome O subunit 1)
XE (Oxidase BO(3) subunit 1) (Cytochrome O ubiquinol oxidase subunit 1).
YN CYOB OR BUSG455.
XS Buchnera aphidicola (subsp. Schizaphis graminum).
XC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
XC Enterobacteriaceae; Buchnera.
XN NCBI_TaxID=98794;
XN [1]
XN SEQUENCE FROM N.A.
XN MEDLINE=22084549; PubMed=12089438;
XA Tamás I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
XA Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
XA "50 million years of genomic stasis in endosymbiotic bacteria.";
XA Science 296:2378-2379(2002).
XC -!- FUNCTION: Cytochrome O terminal oxidase complex is the component
XC of the aerobic respiratory chain that predominates when cells are
XC grown at high aeration. This ubiquinol oxidase shows proton pump
XC activity across the membrane in addition to the electron transfer
XC (by similarity).
XC -!- CATALYTIC ACTIVITY: Ubiquinol-8 + O(2) = Ubiquinone-8 + H(2)O.
XC -!- COFACTOR: Contains two protoheme IX (heme B55 and B562) and
XC copper B (By similarity).
XC -!- PATHWAY: Ubiquinol oxidase catalyzes the terminal step in the
XC electron transport chain.
XC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
XC -!- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
XC
XC This SWISS-PROT entry is copyright. It is produced through a collaboration
XC between the Swiss Institute of Bioinformatics and the EMBL outstation
XC the European Bioinformatics Institute. There are no restrictions on its
XC use by non-profit institutions as long as its content is in no way
XC modified and this statement is not removed. Usage by and for commercial
XC entities requires a license agreement (See http://www.isb-sib.ch/announce/
XC or send an email to license@isb-sib.ch).
XC
XC EMBL; AB014121; AM67998.1; -.
XC InterPro; IPR000883; COX1.
XC Pfam; PF00115; COX1; 1.
XC PRINTS; PR01165; CYCOXIDASE1.
XC PROSITE; PS00077; COX1; 1.
XC Oxidoreductase; Respiratory chain; Transmembrane; Heme; Copper;
CX Hydrogen ion transport; Complete proteome.
CX DOMAIN 1 14 EXTRACELLULAR (POTENTIAL).
CX TRANSMEM 15 35 POTENTIAL.
CX DOMAIN 36 58 CYTOPLASMIC (POTENTIAL).
CX TRANSMEM 59 79 POTENTIAL.
CX DOMAIN 80 106 EXTRACELLULAR (POTENTIAL).
CX TRANSMEM 107 127 POTENTIAL.
CX DOMAIN 128 145 CYTOPLASMIC (POTENTIAL).
CX TRANSMEM 146 166 POTENTIAL.
CX DOMAIN 167 189 EXTRACELLULAR (POTENTIAL).
CX TRANSMEM 190 210 POTENTIAL.
CX DOMAIN 211 225 CYTOPLASMIC (POTENTIAL).
CX TRANSMEM 226 246 POTENTIAL.
CX DOMAIN 247 277 EXTRACELLULAR (POTENTIAL).
CX TRANSMEM 278 298 POTENTIAL.
CX DOMAIN 299 309 CYTOPLASMIC (POTENTIAL).
CX TRANSMEM 310 330 POTENTIAL.
CX DOMAIN 331 347 EXTRACELLULAR (POTENTIAL).
CX TRANSMEM 348 368 POTENTIAL.
CX DOMAIN 369 380 CYTOPLASMIC (POTENTIAL).

```

```

FT TRANSMEM 381 401
FT DOMAIN 402 413
FT TRANSMEM 414 434
FT DOMAIN 435 456
FT TRANSMEM 457 477
FT DOMAIN 478 490
FT TRANSMEM 491 511
FT DOMAIN 512 580
FT TRANSMEM 581 601
FT DOMAIN 602 605
FT TRANSMEM 606 626
FT DOMAIN 627 659
FT METAL 106 106
FT METAL 284 284
FT METAL 288 288
FT METAL 333 333
FT METAL 334 334
FT METAL 419 419
FT METAL 421 421
FT CROSSLINK 284 288
FT SEQUENCE 659 AA; 75028 MW; 23D6FB4B04732D23 CRC64;
SQ
Query Match 8.4%; Score 80; DB 1; Length 659;
Best Local Similarity 20.9%; Pred. No. 3.6;
Matches 40; Conservative 23; Mismatches 54; Indels 74; Gaps 12;
QY 1 MMTWNNALIVFTTV-----VMEVVAALAHK-----YIMHGKG-----WGWHLSH 40
DB 278 LMTWGHPEVYLLPVFGVFSEVATPSKKRLFGYSLVWATLAILTSFIYW-----LH 333
QY 41 HEPRKGA-FEVDNDLYAVVPAIVSIALIYFGSTGIPQWIGAGMTAYGLLYPMVHDLVH 99
DB 334 EFPTMGAGSNVNAFFGITTIAL-----PTGVKLFNWL-----FTWYQGRVH 376
QY 100 QR-----WPFYR-----IPRKYIL--KRLYNAMRHMAVRKE--GCVS-- 134
DB 377 MHSSMLATIGLITFSIGGMTGVLLSIPPADFILHNSFLVAHFHNVIIIGVVYFGCFAGI 436
QY 135 -----FGFL 138
DB 437 NYWPKLFGFI 447

```

Search completed: February 29, 2004, 14:45:17  
Job time : 5.94944 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 4.87104 Seconds  
(without alignments)  
3455.835 Million cell updates/sec

title: US-09-941-947A-36

effect score: 956

sequence: 1 MLWLNALIVFTVVGMEV.....ARSGAARDQGVDTSSGX 175

scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

sarched: 283366 seqs, 96191526 residues

total number of hits satisfying chosen parameters: 283366

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

database: PIR\_78:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	892	93.3	175	2 F37802	criz protein - Erwinia
2	654	68.4	176	2 S52982	beta-carotene hydr
3	271	28.3	151	2 H30469	beta-carotene hydr
4	253	26.5	310	2 T09562	beta-carotene hydr
5	101.5	10.6	383	1 A44227	omega-3 fatty acid
6	99.5	10.4	377	1 J02337	omega-3 fatty acid
7	98.5	10.3	410	2 S75695	hypothetical prote
8	93.5	9.8	386	1 J02335	omega-3 fatty acid
9	93	9.7	418	2 J07872	stearyl-CoA 9-des
10	91	9.5	304	2 A2615	conserved hypotet
11	91	9.5	313	2 E97397	hypothetical prote
12	88	9.2	301	2 G83556	hypothetical prote
13	87	9.1	907	2 S72765	phosphoenolpyruvat
14	87	9.1	934	2 B86981	probable phosphoen
15	84.5	8.8	162	2 A3586	hypothetical membr
16	84.5	8.8	404	2 P00812	omega-3 fatty acid
17	84.5	8.8	455	2 T44519	putrescine/ornithi
18	83.5	8.7	172	2 T11298	NADH2 dehydrogenas
19	83	8.7	350	2 D39364	GDF-1 embryonic gr
20	82.5	8.6	441	2 T03029	omega-3 fatty acid
21	82.5	8.6	446	1 J02336	omega-3 fatty acid
22	82	8.6	662	2 H84984	bo-type ubiquinol
23	81.5	8.5	460	2 T10063	omega-3 fatty acid
24	79.5	8.3	567	2 A33023	potassium-transpor
25	79.5	8.3	567	2 H38261	hypothetical prote
26	78.5	8.2	431	2 T07685	omega-3 fatty acid
27	78	8.2	415	2 AG2044	hypothetical prote
28	78	8.2	551	2 A87019	probable cytochrom
29	78	8.2	751	2 A81816	nitric-oxide reduc

30 78 8.2 751 2 D81062 nitric oxide reduc  
31 77.5 8.1 228 2 AB3510 molybdenum transpo  
32 77.5 8.1 316 2 T19435 hypothetical prote  
33 77.5 8.1 410 2 A70648 probable NADH2 de  
34 77.5 8.1 489 2 S84229 DNA damage-inducib  
35 77 8.1 222 1 D69991 conserved hypotet  
36 77 8.1 3473 1 A46112 genome polyprotein  
37 77 8.1 3473 2 S27927 polyprotein - rice  
38 76.5 8.0 332 1 Q08E39 BGLF3 protein - hu  
39 76.5 8.0 345 2 S72490 N-acetyl-gamma-glu  
40 76.5 8.0 453 1 J02339 omega-3 fatty acid  
41 76.5 8.0 651 2 S44257 phosphotransferase  
42 76 7.9 156 2 S08427 12K protein mlai2A  
43 76 7.9 208 2 T23328 hypothetical prote  
44 76 7.9 300 2 H87631 integral membrane  
45 76 7.9 542 2 C70732 probable integral

#### ALIGNMENTS

##### RESULT 1

F37802

criz protein - Erwinia uredovora

C:Species: Erwinia uredovora

C:Date: 31-May-1991 #sequence\_revision 31-May-1991 #text\_change 20-Jun-2000

C:Accession: F37802

R: Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashim

J. Bacteriol. 172, 6704-6712, 1990

A:Title: Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functio

A:Reference number: A37802; MUID:91072214; PMID:2254247

A:Accession: F37802

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-175 <MIS>

A:Cross-references: GB:D90087; NID:g216681; PIDN:BAAL4129.1; PID:g216687

Query Match 93.3%; Score 892; DB 2; Length 175;

Best Local Similarity 91.4%; Pred. No. 1,8e-81;

Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 MLWLNALIVFTVVGMEVVAALAHKYMHGNGWGLSHHEPRKGAPEVNDLVAVPAI 60

DB 1 MLWLNALIVFTVVGMEVVAALAHKYMHGNGWGLSHHEPRKGAPEVNDLVAVPA 60

QY 61 VSIALTYGSGTGIPLOMTAGMTAYGLLYPMVHDLGVHQRWPERYIPKGYLKLVMNAH 120

DB 61 LSILLIYLGSTGMPLQWPLQWIGAGMTAYGLLYPMVHDLGVHQRWPERYIPKGYLKLVMNAH 120

QY 121 RMHHAVRKGGCVSGFLYAPPLSKLQATLRERHAARSGAARDEQGVDTSSSGX 175

DB 121 RMHHAVRKGGCVSGFLYAPPLSKLQATLRERHAARSGAARDEQGVDTSSSGX 175

##### RESULT 2

S52982

beta-carotene hydroxylase - Erwinia herbicola

C:Species: Erwinia herbicola

C:Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 08-Oct-1999

C:Accession: S52982

R:Hundle, B.; Alberti, M.; Niveststein, V.; Beyer, P.; Kleinig, H.; Armstrong, G.A.; Bur

Mol. Gen. Genet. 245, 406-416, 1994

A:Title: Functional assignment of Erwinia herbicola Ehc 10 carotenoid genes expressed in

A:Reference number: S52976; MUID:95107236; PMID:7808389

A:Accession: S52982

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-176 <HUN>

A:Cross-references: EMBL:M87280; NID:gl48404; PIDN:AAA64983.1; PID:gl48414

Query Match 68.4%; Score 654; DB 2; Length 176;

Best Local Similarity 68.2%; Pred. No. 9e-58;

Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;



45 RMSVVDIIFAVV--ALAVAVYFDSFFFLYWAAG--TLFWAIFVLGHDCGHSFSD 101  
96 -----GLVHQRPPIPRKGYLRLYMAHRMHAHVG 128  
102 IPLINTAVGHILHSFILVPYHGW-----RISHRTHQNHG 136

RESULT 7  
15695  
Prothetical protein sll1376 - *Synechocystis* sp. (strain PCC 6803)  
Species: *Synechocystis* sp.  
Variety: PCC 6803  
Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
Accession: S75695  
Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
IA Res. 3, 109-136, 1996  
Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
Reference number: S74322; MUID:97061201; PMID:8905231  
Accession: S75695  
Status: nucleic acid sequence not shown; translation not shown  
Molecule type: DNA  
Residues: 1-410 <RAN>  
Cross-references: EMBL:D90912; GB:AB001339; NID:gl653228; PIDN:BAA18256.1; PID:d101898  
Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 10.3%; Score 98.5; DB 2; Length 410;  
Best Local Similarity 22.6%; Pred. No. 0.044;  
Matches 51; Conservative 27; Mismatches 51; Indels 97; Gaps 14;  
1 MLIWNNALIVFVVGVMEVA-----ALAH-----XYIMHGNGWGHLSHHPKGAPEV-- 50  
||| : : : : : ||| : : : : :  
3 MAWLMGLGALASVLWELVRDCVHALARWSPLYRLKG-----HHRVFRSLDLSWS 55  
||| : : : : : ||| : : : : :  
51 NDLYA-----VVAIVSIALIYFGSGIWP--LQWIGACWTAY-----CLLY-- 90  
||| : : : : : ||| : : : : :  
56 TEIYQKAIWYNDVPERLWLAFL-----GIWPPPLTWKQFSQWPLILASAGWTILG 109  
||| : : : : : ||| : : : : :  
91 FMVH-----DGLVHQRPPIPRKGYLRLYMAHRMH-----124  
||| : : : : : ||| : : : : :  
110 FLISALARGVGLPNADETLTHRPGPELTPPAWVWVRY--HWRHFDPPNAYFGTL 167  
||| : : : : : ||| : : : : :  
125 -----AVRKGSCVS-----FGLVAPPLSKLOATLRBH 154  
||| : : : : : ||| : : : : :  
168 TLVDXMLGTLSLKGGKIAVTGASGGFG-----QALLQSLH 203  
||| : : : : : ||| : : : : :

RESULT 8  
22335  
omega-3 fatty acid desaturase (EC 1.14.99.-) CF3 [similarity] - *Arabidopsis thaliana*  
Alternate names: protein F23F1.10  
Species: *Arabidopsis thaliana* (mouse-ear cress)  
Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 02-Feb-2001  
Accession: JQ2335; JQ2487; A84703  
Yadav, N.S.; Wierzbicki, A.; Aegerter, M.; Caster, C.S.; Perez-Grau, L.; Kinney, A.J.;  
Russell, S.H.; Feldmann, K.A.; Pierce, J.; Browne, J.  
Ant. Physiol. 103, 467-476, 1993  
Title: Cloning of higher plant omega-3 fatty acid desaturases.  
Reference number: JQ2335; MUID:94302147; PMID:8029334  
Accession: JQ2335  
Molecule type: mRNA  
Residues: 1-386 <YAD>  
Cross-references: GB:D17579; NID:g1030693; PIDN:BAA04505.1; PID:g471091  
Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
mitted to the EMBL Data Library, August 1998  
Description: *Arabidopsis thaliana* chromosome II BAC F23F1 genomic sequence.  
Reference number: Z14675  
Accession: T02487  
Status: translated from GB/EMBL/DBJ  
Molecule type: DNA  
Residues: 1-386 <ROU>

A;Cross-references: EMBL:AC004680; NID:g3420043; PIDN:AAC31854.1; PID:g3420053  
A;Experimental source: cultivar Columbia  
R;lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: A84703  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-386 <STO>  
A;Cross-references: GB:AB002093; NID:g3420053; PIDN:AAC31854.1; GSPDB:GM00139  
C;Comment: The omega-6 and omega-3 fatty acid desaturases introduce the second and the t  
erms of plant membranes.  
C;Genetics:  
A;Gene: At2g29980; F23F1.10  
A;Map position: 2  
A;Introns: 103/2; 133/2; 155/3; 186/3; 248/3; 275/3; 321/3  
C;Superfamily: omega-3 fatty acid desaturase  
C;Keywords: oxidoreductase

Query Match 9.8%; Score 93.5; DB 1; Length 386;  
Best Local Similarity 26.5%; Pred. No. 0.13;  
Matches 27; Conservative 17; Mismatches 31; Indels 27; Gaps 5;  
QY 44 RKAPEVNDLYAVWPAIVSIALIYFGSGIWPLOWIGACWTAYGLLYPMVHD----- 95  
||| : : : : : ||| : : : : :  
Db 54 RMSVYVVDIIAV--ALATAAVYVDSWFLWFLYWAAG--TLFWAIFVLGHDCGHSFSD 110  
||| : : : : : ||| : : : : :  
QY 96 -----GLVHQRPPIPRKGYLRLYMAHRMHAHVG 128  
||| : : : : : ||| : : : : :  
Db 111 IPLLNSVVGHILHS-----FILVPYHGW-----RISHRTHQNHG 145  
||| : : : : : ||| : : : : :

RESULT 9  
JC7872  
stearoyl-CoA 9-desaturase (EC 1.14.19.1), FAD3 - *Chlorella vulgaris*  
N;Alternate names: acyl-CoA desaturase; delta9-desaturase; omega-3 stearoyl-CoA desatura  
C;Species: *Chlorella vulgaris*  
C;Date: 09-Dec-2002 #sequence\_revision 09-Dec-2002 #text\_change 31-Mar-2003  
C;Accession: JC7872  
R;Suga, K.; Honjo, K.; Furuya, N.; Shimizu, H.; Nishi, K.; Shinohara, F.; Hirabaru, Y.;  
Biosci. Biotechnol. Biochem. 66, 1314-1327, 2002  
A;Title: Two low-temperature-inducible *Chlorella* genes for delta12 and omega-3 fatty aci  
s cerevisiae, and expression of omega-3 fed in *Nicotiana tabacum*.  
A;Reference number: JC7871; MUID:22152188; PMID:12162554  
A;Accession: JC7872  
A;Molecule type: mRNA  
A;Residues: 1-418 <SUG>  
A;Cross-references: DBJ:AB075527  
C;Comment: This enzyme is involved in low temperature adaptation and the development of  
C;Genetics:  
A;Gene: fad3  
C;Keywords: oxidoreductase

Query Match 9.7%; Score 93; DB 2; Length 418;  
Best Local Similarity 23.1%; Pred. No. 0.16;  
Matches 36; Conservative 19; Mismatches 53; Indels 48; Gaps 6;  
QY 57 VPAIVSIALIYFGSGIWPLOWIGACWTAYGLLYPMVHDGLVHQRPPIPRKGYLRLYMAHRMH 105  
||| : : : : : ||| : : : : :  
Db 102 IVAALAIAGYITGNPLVWFLYFLQG--TMFWALFVVGHD--CGHQSWSNNKTLNDFVGNIV 159  
||| : : : : : ||| : : : : :  
QY 106 ----YIPKGYLRLYMAHRMHAHVG-----KEGCVSFGFL 138  
||| : : : : : ||| : : : : :  
Db 160 HSIWVPHGW-----RISHRTHAHGHVENDSWHPVKSNEYKLDKWLKGLLPPF- 214  
||| : : : : : ||| : : : : :  
QY 139 YAPPLSKLOATLRHRAARSAAARDEQGVDTSSG 174  
||| : : : : : ||| : : : : :  
Db 215 ---PLFAYPFVLLNRSCKGSHYDPSKSLFTASEG 247  
||| : : : : : ||| : : : : :

[illegible]

```

A; Gene: pbc
C; Superfamily: phosphoenolpyruvate carboxylase
C; Keywords: carbon-carbon lyase; carboxy-lyase

Query Match          9.1%; Score 87; DB 2; Length 907;
Best Local Similarity 23.5%; Pred. NO. 1.4;
Matches 38; Conservative 21; Mismatches 53; Indels 50; Gaps 7;

Qy 44  RKGAEEVNDLYAVFAIV--SIALITYFGSGIWFPLQWIGAGMTAYGLLYFWHDGLVH 99
      ||| : : : : ||| : : : : ||| : : : : ||| : : : :
Db 725 RPTTSIADLRAPWLAWSQSRVPLPWGYGTGSAPQOWVAAGPSESORVENLHD--LY 782
      ||| : : : : ||| : : : : ||| : : : : ||| : : : :

Qy 100 QRWPF-----RYP-----RKGYLRLYMAHR-----MHAVR 127

```

|||||  
783 QRWPFPRSVLSNMAQVLAKSDGLAARYAELVWDEALRRRVEFKIADERRRTIAIHKLIT 842  
128 GKGCVSFG-----FLYAPPLSKLQATLRERHAARSG 159  
843 GHDDLLADNPALARSVFNRPFPYLEPLNHLQVELLRRY--RSG 882

SULT 14  
6981  
obable phosphoenolpyruvate carboxylase [imported] - Mycobacterium leprae  
Species: Mycobacterium leprae  
Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
Accession: B86981  
Cole, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho  
; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,  
m, M.A.; Rutherford, K.M.  
ture 409, 1007-1011, 2001  
Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq  
Title: Massive gene decay in the leprosy bacillus.  
Reference number: A86909; MUID:21128732; PMID:11234002  
Accession: B86981  
Status: preliminary  
Molecule type: DNA  
Residues: 1-934 <STO>  
Cross-references: GB:AL450380; NID:gl3092776; PIDN:CAC30086.1; GSPDB:GN00147  
Genetics:  
Gene: ppc  
Superfamily: phosphoenolpyruvate carboxylase

Query Match 9.1%; Score 87; DB 2; Length 934;  
Best Local Similarity 23.5%; Pred.No. 1.5;  
Matches 38; Conservative 21; Mismatches 53; Indels 50; Gaps 7;  
44 RKGAPEVNDLYAVFAIV-----SIALYFGSTGIWPLQWIGAGMTAYGLLYFMVHEDGLVH 99  
752 RKPTTSIADLRAPWVLANSQSRWMLPGWYGTGSAFQWVAAPESQSRVEMLED--LY 809  
100 QRWPF-----RYIP-----RKGYLKLRYMAHR-----MHHAVR 127  
810 QRWPFPRSVLSNMAQVLAKSDGLAARYAELVWDEALRRRVEFKIADERRRTIAIHKLIT 869

128 GKGCVSFG-----FLYAPPLSKLQATLRERHAARSG 159  
870 GHDDLLADNPALARSVFNRPFPYLEPLNHLQVELLRRY--RSG 909

SULT 15  
3586  
pochetical membrane spanning protein BMEII0616 [imported] - Brucella melitensis (strai  
Species: Brucella melitensis  
Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
Accession: AG3586  
DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,  
Mazur, M.; Gotsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
oc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
Title: The genome sequence of the facultative intracellular pathogen Brucella melitens  
Reference number: AD3252; PMID:11756688  
Accession: AG3586  
Status: preliminary  
Molecule type: DNA  
Residues: 1-162 <KUR>  
Cross-references: GB:AE008918; PIDN:AAL53858.1; PID:gi7984795; GSPDB:GN00191  
Experimental source: strain 16M  
Genetics:  
Gene: BMEII0616  
Map position: II

Query Match 8.8%; Score 84.5; DB 2; Length 162;  
Best Local Similarity 27.0%; Pred.No. 0.41;  
Matches 41; Conservative 23; Mismatches 41; Indels 47; Gaps 13;  
4 INKALIVFV--TVVGMEVVAALAHKYNHGG-----WG-----WGMHLSR-----HE 42

|||||  
5 IWRGILIGIGATVV-MDIWAQLL---ALLPGQSRPKNGLVGRWFHLEHGOIFHDDISQSE 61  
43 PRK-----GAFVNDLYAVFAIVSIALYFGSTGIWPLQWIGAGMTAYGLL----- 89  
62 PCRHEVALGWTGHTYAVGILYGVIFALYGGSA-AWFANPFLP-AWI-----FGILTTAAG 113  
90 YFMVHVDGLVHQRWPFYIPRKGYLKLRL-YMAH 120  
114 WFLQPG-L-GIGWAASKLPNAGNVRILNLIAH 144

Search completed: February 29, 2004, 14:52:57  
Job time : 7.87104 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

[ protein - protein search, using sw model

in on: February 29, 2004, 14:51:24 ; Search time 11.619 Seconds  
(without alignments)  
3180.293 Million cell updates/sec

file: US-09-941-947A-36

Effect score: 956  
Sequence: 1 MLMTWNLALIVFTVVGMEVV.....ARSGAARDEGGVDTSSGK 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Aligned: 809742 seqs, 211153259 residues

Initial number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	956	100.0	175	10	US-09-941-947A-36
2	956	100.0	175	14	US-10-115-571A-12
3	892	93.3	175	9	US-09-323-998D-7
4	654	68.4	175	9	US-09-323-998D-6
5	499	52.2	162	9	US-09-323-998D-5
6	495	51.8	162	10	US-09-920-923-34
7	488	51.0	162	9	US-09-323-998D-8
8	487	50.9	169	9	US-09-547-267-9
9	487	50.9	169	14	US-10-166-225A-193
10	487	50.9	169	9	US-09-323-998D-4
11	247	25.8	294	9	US-09-323-998D-1
12	245	25.6	309	15	US-10-259-194A-398
13	101.5	10.6	383	14	US-10-115-571A-3
14	101.5	10.6	383	14	US-10-115-571A-10
15	100.5	10.5	380	14	US-10-115-571A-1

16	99.5	10.4	377	9	US-09-837-751-34	Sequence 34, Appl
17	99.5	10.4	377	14	US-10-115-571A-2	Sequence 2, Appli
18	99.5	10.4	377	14	US-10-115-571A-9	Sequence 9, Appli
19	93.5	8.8	386	14	US-10-115-571A-4	Sequence 4, Appli
20	93.5	9.8	386	14	US-10-115-571A-11	Sequence 11, Appl
21	92	9.6	379	14	US-10-115-571A-68	Sequence 68, Appl
22	91.5	9.6	358	14	US-10-115-571A-16	Sequence 16, Appl
23	90.5	9.5	549	15	US-10-369-493-3466	Sequence 3466, Ap
24	87	9.1	335	14	US-10-156-761-11607	Sequence 11607, A
25	86.5	9.0	351	14	US-10-115-571A-15	Sequence 15, Appl
26	86.5	9.0	435	15	US-10-310-154-468	Sequence 468, App
27	86	9.0	352	14	US-10-115-571A-19	Sequence 19, Appl
28	84.5	8.8	360	14	US-10-115-571A-26	Sequence 26, Appl
29	84	8.8	375	14	US-10-115-571A-14	Sequence 14, Appl
30	83	8.7	350	15	US-10-295-027-246	Sequence 246, App
31	83	8.7	413	15	US-10-369-493-3871	Sequence 3871, Ap
32	82.5	8.6	362	14	US-10-115-571A-18	Sequence 18, Appl
33	82.5	8.6	363	14	US-10-115-571A-20	Sequence 20, Appl
34	82.5	8.6	369	14	US-10-115-571A-21	Sequence 21, Appl
35	82.5	8.6	382	14	US-10-115-571A-69	Sequence 69, Appl
36	82.5	8.6	446	15	US-10-310-154-467	Sequence 467, App
37	81.5	8.5	321	14	US-10-115-571A-37	Sequence 37, Appl
38	81.5	8.5	377	14	US-10-115-571A-17	Sequence 17, Appl
39	81.5	8.5	429	15	US-10-369-493-3762	Sequence 3762, Ap
40	81	8.5	347	15	US-10-259-194A-160	Sequence 160, App
41	81	8.5	368	14	US-10-115-571A-23	Sequence 23, Appl
42	80	8.4	371	9	US-09-885-188-10	Sequence 10, Appl
43	80	8.4	372	9	US-09-885-189-10	Sequence 10, Appl
44	79	8.3	269	14	US-10-156-761-14555	Sequence 14555, A
45	78.5	8.2	362	14	US-10-115-571A-30	Sequence 30, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-941-947A-36  
; Sequence 36, Application US/09941947A  
; Publication No. US20030003528A1  
; GENERAL INFORMATION:  
; APPLICANT: Brzostowicz, Patricia C.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: Dicosimo, Deana J.  
; APPLICANT: Koffas, Mattheos  
; APPLICANT: Miller, Edward S. Jr.  
; APPLICANT: Odom, J. Martin  
; APPLICANT: Picataggio, Steve  
; APPLICANT: Rouviere, Pierre E.  
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE  
; FILE REFERENCE: CL1903 US NA  
; CURRENT APPLICATION NUMBER: US/09/941.947A  
; CURRENT FILING DATE: 2001-09-01  
; PRIOR APPLICATION NUMBER: 60/229,907  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,858  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 36  
; LENGTH: 175  
; TYPE: PRT  
; ORGANISM: Pantoea stewartii  
US-09-941-947A-36

Query Match 100.0%; Score 956; DB 10; Length 175;  
Best Local Similarity 100.0%; Pred. No. 9e-101;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLMTWNLALIVFTVVGMEVVAALAHKYMCGWGHLSHHEPRKGAPEVNDLYAVVFAI 60  
DB 1 MLMTWNLALIVFTVVGMEVVAALAHKYMCGWGHLSHHEPRKGAPEVNDLYAVVFAI 60  
QY 61 VSIALIVFGSTGIWPLQWIGAGMTAYGLLYFMVDGLVHQWPFYIPRKGYLRLTMH 120



b 61 VSTALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 120  
y 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEODGVDTSSSGK 175  
b 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEODGVDTSSSGK 175

## RESULT 2

S-10-218-118-12  
Sequence 12, Application US/10218118  
Publication No. US2003014831A1  
GENERAL INFORMATION:  
APPLICANT: Byrostowicz, Patricia  
APPLICANT: Rouviere, Pierre  
APPLICANT: Picataggio, Stephen  
APPLICANT: Cheng, Qiong  
TITLE OF INVENTION: Genes Encoding Carotenoid Compounds  
FILE REFERENCE: CL1876 US NA  
CURRENT APPLICATION NUMBER: US/10/218,118  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: 60/312,646  
PRIOR FILING DATE: 2001-08-15  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 12  
LENGTH: 175  
TYPE: PRT  
ORGANISM: Pantoea stewartii  
S-10-218-118-12

Query Match 100.0%; Score 956; DB 14; Length 175;  
Best Local Similarity 100.0%; Pred. No. 9e-101;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
y 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 60  
b 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 60  
y 61 VSTALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 120  
b 61 VSTALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 120  
y 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEODGVDTSSSGK 175  
b 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEODGVDTSSSGK 175

## RESULT 3

S-09-323-998D-7  
Sequence 7, Application US/09323998D  
Patent No. US20020102631A1  
GENERAL INFORMATION:  
APPLICANT: CUNNINGHAM JR., FRANCIS X.  
APPLICANT: SUN, ZAIREN  
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND  
METHODS OF USE THEREOF  
FILE REFERENCE: 108172-09019  
CURRENT APPLICATION NUMBER: US/09/323,998D  
CURRENT FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: 09/088,724  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 09/088,725  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 08/937,155  
PRIOR FILING DATE: 1997-09-25  
PRIOR APPLICATION NUMBER: 08/624,125  
PRIOR FILING DATE: 1996-03-29  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 175  
TYPE: PRT  
ORGANISM: Pantoea stewartii

; ORGANISM: Erwinia uredovora  
US-09-323-998D-7  
Query Match 93.3%; Score 892; DB 9; Length 175;  
Best Local Similarity 91.4%; Pred. No. 1.8e-93;  
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 60  
Db 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 60  
Qy 61 VSTALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 120  
Db 61 VSTALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 120  
Qy 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEODGVDTSSSGK 175  
Db 121 RMEHAVRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEODGVDTSSSGK 175

## RESULT 4

US-09-323-998D-6  
Sequence 6, Application US/09323998D  
Patent No. US20020102631A1  
GENERAL INFORMATION:  
APPLICANT: CUNNINGHAM JR., FRANCIS X.  
APPLICANT: SUN, ZAIREN  
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND  
METHODS OF USE THEREOF  
FILE REFERENCE: 108172-09019  
CURRENT APPLICATION NUMBER: US/09/323,998D  
CURRENT FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: 09/088,724  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 09/088,725  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 08/937,155  
PRIOR FILING DATE: 1997-09-25  
PRIOR APPLICATION NUMBER: 08/624,125  
PRIOR FILING DATE: 1996-03-29  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 175  
TYPE: PRT  
ORGANISM: Erwinia herbicola  
US-09-323-998D-6

Query Match 68.4%; Score 654; DB 9; Length 175;  
Best Local Similarity 68.2%; Pred. No. 2.3e-66;  
Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;  
Qy 6 NALIVFTVVGMEVVAALAHKYIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 65  
Db 3 NSLIVLSVTAMEGIAAFTHRVIMHGCGWGHLSHHEPRKGAFVNDLYAVVFAI 62  
Qy 66 IYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 125  
Db 63 IYGTAGVWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPFYIPRKGYLKELYMAH 122  
Qy 126 VRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEODGVDTSSSGK 173  
Db 123 VRGKGCVSFGFLYAPPLSKLOATLRERHARSGAARDEODGVDTSSSGK 172

## RESULT 5

US-09-323-998D-5  
Sequence 5, Application US/09323998D  
Patent No. US20020102631A1  
GENERAL INFORMATION:  
APPLICANT: CUNNINGHAM JR., FRANCIS X.  
APPLICANT: SUN, ZAIREN  
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND  
METHODS OF USE THEREOF

## TITLE OF INVENTION: METHODS OF USE THEREOF

FILE REFERENCE: 108172-09019  
CURRENT APPLICATION NUMBER: US/09/323,998D  
CURRENT FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: 09/088,724  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 09/088,725  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 08/937,155  
PRIOR FILING DATE: 1997-09-25  
PRIOR APPLICATION NUMBER: 08/624,125  
PRIOR FILING DATE: 1996-03-29  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 162  
TYPE: PRT  
ORGANISM: Alicycobacteres sp.  
US-09-323-998D-5

Query Match 52.2%; Score 499; DB 9; Length 162;  
Best Local Similarity 57.8%; Pred. No. 9.6e-49;  
Matches 93; Conservative 22; Mismatches 38; Indels 8; Gaps 4;

8 LIVFVTVGMEVVAALAHKYIMHG-WGCGWHLSHHEPRKGAPEVNDLYAVVFAIVSIA 66  
5 LIVVATVLMVELTAYSVHRWIMHGPGWGMKSHHEHDHLEKNDLYGVVFAVIA 64

67 YFGSTGIMP-LQWICAGMTAYGLLYFVMDGLVHQRPFRYIPRKGYLKRLYMAHARM 125  
65 TVGAVW-WPVLWNLWIALGTVVGLIYFVLDGLVHQRPFRYIPRKGYLKRLYMAHARM 123

126 VRGEGCVSFGFLYAPPLSKLOATLRERHAARSGAARDEQD 166  
124 VEGRDHCVSFGFIYAPPDVCLKQDLK-----RSGVLRPQDE 159

## RESULT 6

US-09-920-923-34  
Publication No. US2003002273A1  
GENERAL INFORMATION:  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tsygankov, Yuri  
TITLE OF INVENTION: Improved Fermentative Carotenoid Production  
FILE REFERENCE: Improved Fermentative Carotenoid  
CURRENT APPLICATION NUMBER: US/09/920,923  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 08/980,832  
PRIOR FILING DATE: 1997-12-01  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 34  
LENGTH: 162  
TYPE: PRT  
ORGANISM: E-396  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Unknown  
US-09-920-923-34

Query Match 51.8%; Score 495; DB 10; Length 162;  
Best Local Similarity 58.0%; Pred. No. 2.7e-48;  
Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;

6 NALIVFTVVGMEVVAALAHKYIMHG-WGCGWHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64  
3 NFLIVVATVLMVELTAYSVHRWIMHGPGWGMKSHHEHDHLEKNDLYGLVFAVIATV 62

65 LIYFGSTGIMP-LQWICAGMTAYGLLYFVMDGLVHQRPFRYIPRKGYLKRLYMAHARM 122  
63 LFTVG--WVAPVLMWIALGTVVGLIYFVLDGLVHQRPFRYIPRKGYLKRLYMAHARM 120

123 HHAVRGEGCVSFGFLYAPPLSKLOATLRERHAARSGAARDE 164

Db 121 HHAVEGRDHCVSFGFIYAPPDVCLKQDLK-----TSGVLRPQDE 157

## RESULT 7

US-09-323-998D-8  
Sequence 8, Application US/09323998D  
Patent No. US20020102631A1  
GENERAL INFORMATION:  
APPLICANT: CUNNINGHAM JR., FRANCIS X.  
APPLICANT: SUN, ZAIREN  
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND  
FILE REFERENCE: 108172-09019  
CURRENT APPLICATION NUMBER: US/09/323,998D  
CURRENT FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: 09/088,724  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 09/088,725  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 08/937,155  
PRIOR FILING DATE: 1997-09-25  
PRIOR APPLICATION NUMBER: 08/624,125  
PRIOR FILING DATE: 1996-03-29  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 162  
TYPE: PRT  
ORGANISM: Agrobacterium auranticum  
US-09-323-998D-8

Query Match 51.0%; Score 488; DB 9; Length 162;  
Best Local Similarity 57.4%; Pred. No. 1.7e-47;  
Matches 93; Conservative 19; Mismatches 40; Indels 10; Gaps 4;

6 NALIVFTVVGMEVVAALAHKYIMHG-WGCGWHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64  
3 NFLIVVATVLMVELTAYSVHRWIMHGPGWGMKSHHEHDHLEKNDLYGLVFAVIATV 62

65 LIYFGSTGIMP-LQWICAGMTAYGLLYFVMDGLVHQRPFRYIPRKGYLKRLYMAHARM 122  
63 LFTVG--WVAPVLMWIALGTVVGLIYFVLDGLVHQRPFRYIPRKGYLKRLYMAHARM 120

123 HHAVRGEGCVSFGFLYAPPLSKLOATLRERHAARSGAARDE 164  
121 HHAVEGRDHCVSFGFIYAPPDVCLKQDLK-----MSGVLRPQDE 157

## RESULT 8

US-09-547-267-9  
Sequence 9, Application US/09547267  
Patent No. US20020147371A1  
GENERAL INFORMATION:  
APPLICANT: Rohmann, Hans-Peter  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tessier, Michel  
APPLICANT: van Loon, Adolphus  
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: NJ  
COUNTRY: USA  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/547,267  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/660,645  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pokras, Bruce A.  
REGISTRATION NUMBER: 32,748  
REFERENCE/DOCKET NUMBER: RAN 6002/170  
TELEPHONE: (201) 235-5801  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 169 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
S-09-547-267-9

Query Match 50.9%; Score 487; DB 9; Length 169;  
Best Local Similarity 51.5%; Pred. No. 2.3e-47;  
Matches 88; Conservative 31; Mismatches 44; Indels 8; Gaps 4;  
  
Y 5 WNALI-VFVTVGMEVVAALAHKYIMHG-WGNGWHLSSH-EPRKGAFVNDLYAVVFAIV 61  
b 4 WAAILTVILTVAAAMETATYSVHRWIMHGPGWGHKSHHDEHDHDALEKNDLYGVIFAVI 63  
  
Y 62 SIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFPRYIPRKGYLKRLYMAHR 121  
b 64 SIVLFAIGMGSDLAWLAVGVTCYGLIYFLHDGLVHGWRPFPRYVPRKGYLRRVYQAGR 123  
  
Y 122 MHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQDGVDTSS 172  
b 124 MHAVHGRENCVSFGFIWAPSVDSLAKELK-----RSGALLKDREGADNT 169

RESULT 9  
S-09-920-923-6  
Sequence 6, Application US/09920923  
Publication No. US2003002273A1  
GENERAL INFORMATION:  
APPLICANT: Pasaontes, Luis  
APPLICANT: Tsygankov, Yuri  
TITLE OF INVENTION: Improved Fermentative Carotenoid Production  
FILE REFERENCE: Improved Fermentative Carotenoid  
CURRENT APPLICATION NUMBER: US/09/920,923  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 08/980,832  
PRIOR FILING DATE: 1997-12-01  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 6  
LENGTH: 169  
TYPE: PRT  
ORGANISM: *Plavobacterium* sp. R1534  
S-09-920-923-6

Query Match 50.9%; Score 487; DB 10; Length 169;  
Best Local Similarity 51.5%; Pred. No. 2.3e-47;  
Matches 88; Conservative 31; Mismatches 44; Indels 8; Gaps 4;  
  
Y 5 WNALI-VFVTVGMEVVAALAHKYIMHG-WGNGWHLSSH-EPRKGAFVNDLYAVVFAIV 61  
b 4 WAAILTVILTVAAAMETATYSVHRWIMHGPGWGHKSHHDEHDHDALEKNDLYGVIFAVI 63  
  
Y 62 SIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFPRYIPRKGYLKRLYMAHR 121  
b 64 SIVLFAIGMGSDLAWLAVGVTCYGLIYFLHDGLVHGWRPFPRYVPRKGYLRRVYQAGR 123  
  
Y 122 MHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQDGVDTSS 172

Db 124 MHAVHGRENCVSFGFIWAPSVDSLAKELK-----RSGALLKDREGADNT 169  
  
RESULT 10  
US-10-166-225A-183  
Sequence 183, Application US/10166225A  
Publication No. US20030148416A1  
GENERAL INFORMATION:  
APPLICANT: BERRY, Alan  
APPLICANT: BRETZEL, Werner  
APPLICANT: HUMBELIN, Markus  
APPLICANT: LOPEZ-ULIBARRI, Rual  
APPLICANT: MAYER, Anne F.  
APPLICANT: YELISEEV, Alexei A.  
TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION  
FILE REFERENCE: C38435/121966  
CURRENT APPLICATION NUMBER: US/10/166,225A  
CURRENT FILING DATE: 2002-06-05  
NUMBER OF SEQ ID NOS: 197  
SOFTWARE: Patent in version 3.1  
SEQ ID NO 183  
LENGTH: 169  
TYPE: PRT  
ORGANISM: *Paracoccus* sp. R1534  
US-10-166-225A-183

Query Match 50.9%; Score 487; DB 14; Length 169;  
Best Local Similarity 51.5%; Pred. No. 2.3e-47;  
Matches 88; Conservative 31; Mismatches 44; Indels 8; Gaps 4;  
  
Y 5 WNALI-VFVTVGMEVVAALAHKYIMHG-WGNGWHLSSH-EPRKGAFVNDLYAVVFAIV 61  
b 4 WAAILTVILTVAAAMETATYSVHRWIMHGPGWGHKSHHDEHDHDALEKNDLYGVIFAVI 63  
  
Y 62 SIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFPRYIPRKGYLKRLYMAHR 121  
b 64 SIVLFAIGMGSDLAWLAVGVTCYGLIYFLHDGLVHGWRPFPRYVPRKGYLRRVYQAGR 123  
  
Y 122 MHAVRGKGCVSFGFLYAPPLSKLQATLRERHAARSGAARDQDGVDTSS 172  
b 124 MHAVHGRENCVSFGFIWAPSVDSLAKELK-----RSGALLKDREGADNT 169

RESULT 11  
US-09-323-998D-4  
Sequence 4, Application US/09323998D  
Patent No. US20020102631A1  
GENERAL INFORMATION:  
APPLICANT: CUNNINGHAM JR., FRANCIS X.  
APPLICANT: SUN, ZAIREN  
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND METHODS OF USE THEREOF  
FILE REFERENCE: 108172-09019  
CURRENT APPLICATION NUMBER: US/09/323,998D  
CURRENT FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: 09/088,724  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 09/088,725  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 08/937,155  
PRIOR FILING DATE: 1997-09-25  
PRIOR APPLICATION NUMBER: 05/624,125  
PRIOR FILING DATE: 1996-03-29  
NUMBER OF SEQ ID NOS: 61  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 4  
LENGTH: 294  
TYPE: PRT  
ORGANISM: *Arabidopsis thaliana*  
US-09-323-998D-4

Query Match 25.8%; Score 247; DB 9; Length 294;

Best Local Similarity 41.2%; Pred. No. 1e-19;  
Matches 54; Conservative 24; Mismatches 47; Indels 6; Gaps 3;  
15 VGEVVAALAHKYMIGWGNGWHLSSHEPRKGAPEVNDLVAVVFAIVSIALIYFG--STG 72  
131 VGEVFWARWALWHLASLNNWESHKREGPELNDVFAIVNAGPAIGLLSYGFNKG 190  
73 IWLQWAGM--TAYGLLYFMVHDGLVHQRWPFYIPRKGYLKLRYMAHRMHHAVRGKE 130  
191 LVPELCFAGLIGITVFGIAYMFVHDGLVHQRWPFYIPRKGYLKLRYMAHRMHHAVRGKE 248  
131 GCVSFGLFLYP 141  
249 NGVPYGLFLGP 259

SULT 12  
-10-259-194A-398  
Sequence 398, Application US/10259194A  
Publication No. US20040010815A1  
GENERAL INFORMATION:

APPLICANT: Lange, Markus B.  
APPLICANT: Ghasseman, Majid  
APPLICANT: Briggs, Steven P.  
APPLICANT: Cooper, Bret  
APPLICANT: Glazebrook, Jane  
APPLICANT: Goff, Stephen A.  
APPLICANT: Katagiri, Fumiyaki  
APPLICANT: Kreps, Joel  
APPLICANT: Moughamer, Todd  
APPLICANT: Provart, Nicholas  
APPLICANT: Ricke, Darrell  
APPLICANT: Zhu, Tong

TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES

CURRENT APPLICATION NUMBER: US/10/259,194A  
CURRENT FILING DATE: 2003-01-07

PRIOR APPLICATION NUMBER: US 60/325,277

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: US 60/370,743

PRIOR FILING DATE: 2002-04-04

PRIOR APPLICATION NUMBER: US 60/370,620

PRIOR FILING DATE: 2002-04-04

NUMBER OF SEQ ID NOS: 662  
SOFTWARE: Patentlist.pl version 3.0.4 (C) 2001 Syngenta

SEQ ID NO 398

LENGTH: 309

TYPE: PRT

ORGANISM: Oryza sativa

-10-259-194A-398

Query Match 25.6%; Score 245; DB 15; Length 309;  
Best Local Similarity 41.2%; Pred. No. 1.8e-19;  
Matches 54; Conservative 23; Mismatches 48; Indels 6; Gaps 4;

15 VGEVVAALAHKYMIGWGNGWHLSSHEPRKGAPEVNDLVAVVFAIVSIALIYFG--STG 72

158 VGXEPQAWHRLSWHLASLNNWESHKREGPELNDVFAIVNAGPAIGLLSYGFNKG 217

73 IWLQWAGM--TAYGLLYFMVHDGLVHQRWPFYIPRKGYLKLRYMAHRMHHAVRGKE 130

218 IVFGLCFAGLIGITLFGMAYFMVHDGLVHQRWPFYIPRKGYLKLRYMAHRMHHAVRGKE 276

131 GCVSFGLFLYP 141

277 G-VPYGLFLGP 286

SULT 13

-10-115-571A-3

Sequence 3, Application US/10115571A

Publication No. US20030150020A1

GENERAL INFORMATION:

APPLICANT: HER MAJESTY IN RIGHT OF CANADA as represented by THE MINISTER OF  
APPLICANT: AGRICULTURE AND AGRI-FOOD CANADA  
APPLICANT: SOMERS, Daryl  
APPLICANT: RAKOW, Gerhard  
TITLE OF INVENTION: PLANT FATTY ACID DESATURASES AND ALLELES THEREFOR  
FILE REFERENCE: 81601-28  
CURRENT APPLICATION NUMBER: US/10/115,571A  
CURRENT FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: CA 2,284,246  
PRIOR FILING DATE: 1999-10-01  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 383  
TYPE: PRT  
ORGANISM: Brassica napus  
PUBLICATION INFORMATION:  
DATABASE ACCESSION NUMBER: GenBank/L01418  
DATABASE ENTRY DATE: 1993-04-27  
US-10-115-571A-3

Query Match 10.6%; Score 101.5; DB 14; Length 383;  
Best Local Similarity 25.5%; Pred. No. 0.0052;  
Matches 26; Conservative 20; Mismatches 29; Indels 27; Gaps 5;

QY 44 RKGAPEVNDLVAVVFAIVSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHD----- 95

Db 51 RSMGYVTRDIFAV--AALAAAVYFDSWFLWPLYVVAQG-TLFWAIFVLGHDGCHGSPSD 107

QY 96 -----GLVHQRWPFYIPRKGYLKLRYMAHRMHHAVRG 128

Db 108 IPLNSVVGHLHS---FILVPYHGW----RISHRTHQNHG 142

RESULT 14

US-10-115-571A-10

Sequence 10, Application US/10115571A

Publication No. US20030150020A1

GENERAL INFORMATION:

APPLICANT: HER MAJESTY IN RIGHT OF CANADA as represented by THE MINISTER OF

APPLICANT: AGRICULTURE AND AGRI-FOOD CANADA

APPLICANT: SOMERS, Daryl

APPLICANT: RAKOW, Gerhard

TITLE OF INVENTION: PLANT FATTY ACID DESATURASES AND ALLELES THEREFOR

FILE REFERENCE: 81601-28

CURRENT APPLICATION NUMBER: US/10/115,571A

CURRENT FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: CA 2,284,246

PRIOR FILING DATE: 1999-10-01

NUMBER OF SEQ ID NOS: 73

SOFTWARE: PatentIn version 3.1

SEQ ID NO 10

LENGTH: 383

TYPE: PRT

ORGANISM: Brassica napus

PUBLICATION INFORMATION:

DATABASE ACCESSION NUMBER: GenBank/P48624

DATABASE ENTRY DATE: 1996-02-01

RELEVANT RESIDUES: (1) ..(383)

US-10-115-571A-10

Query Match 10.6%; Score 101.5; DB 14; Length 383;  
Best Local Similarity 25.5%; Pred. No. 0.0052;  
Matches 26; Conservative 20; Mismatches 29; Indels 27; Gaps 5;

QY 44 RKGAPEVNDLVAVVFAIVSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHD----- 95

Db 51 RSMGYVTRDIFAV--AALAAAVYFDSWFLWPLYVVAQG-TLFWAIFVLGHDGCHGSPSD 107

QY 96 -----GLVHQRWPFYIPRKGYLKLRYMAHRMHHAVRG 128

Db 108 IPLNSVVGHLHS---FILVPYHGW----RISHRTHQNHG 142

RESULT 15  
S-10-115-571A-1  
Sequence 1, Application US/10115571A  
Publication No. US20030150020A1  
GENERAL INFORMATION:  
APPLICANT: HER MAJESTY IN RIGHT OF CANADA as represented by THE MINISTER OF  
APPLICANT: AGRICULTURE AND AGRI-FOOD CANADA  
APPLICANT: SOMERS, Daryl  
APPLICANT: RAKOW, Gerhard  
TITLE OF INVENTION: PLANT FATTY ACID DESATURASES AND ALLELES THEREFOR  
FILE REFERENCE: 81601-28  
CURRENT APPLICATION NUMBER: US/10/115,571A  
CURRENT FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: CA 2,284,246  
PRIOR FILING DATE: 1999-10-01  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 380  
TYPE: PRT  
ORGANISM: Brassica napus  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (95)..(99)  
OTHER INFORMATION: Histidine box sequence  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (213)..(213)  
OTHER INFORMATION: Position of amino acid substitution in accordance with various as  
OTHER INFORMATION: pects of the invention  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (275)..(275)  
OTHER INFORMATION: Position of amino acid substitution in accordance with various as  
OTHER INFORMATION: pects of the invention  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (347)..(347)  
OTHER INFORMATION: Position of amino acid substitution in accordance with various as  
OTHER INFORMATION: pects of the invention  
S-10-115-571A-1  
Query Match 10.5%; Score 100.5; DB 14; Length 380;  
Best Local Similarity 26.3%; Pred. No. 0.0067;  
Matches 26; Conservative 17; Mismatches 35; Indels 21; Gaps 4;  
ly 44 RKGAFEVNDLYAVVFAIVSIALIYFGSTGIWPLQMGATAYGLLYFMVHD----- 95  
lb 48 RMSYVARDIFSIV--ALAAVAVYFDSEFFWPLYNAAQG-TLFWAIFVLGHDCGHGSPSD 104  
ly 96 -----GLVHQRPFRYIPKGYLKLKLYMAHRMHHAVRG 128  
lb 105 IPLLNTAVGHILHSFILVPYHGW----RMSHRTHQNHG 139  
Search completed: February 29, 2004, 15:28:06  
Ob time : 12.619 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on on: February 29, 2004, 14:35:44 ; Search time 5.31793 Seconds  
(without alignments)  
1698.885 Million cell updates/sec

le: US-09-941-947a-36

irect score: 956

quence: 1 MLMTWNLIVFTVVGMEVV.....ARSGAARDEQDGYDTSSGK 175

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

tal number of hits satisfying chosen parameters: 389414

imum DB seq length: 0

ximum DB seq length: 2500000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Issued Patents AA.\*

- 1: /cgm2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgm2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgm2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgm2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgm2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*
- 6: /cgm2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	892	93.3	175	1	US-07-783-705A-6
2	892	93.3	175	1	US-08-624-125-7
3	892	93.3	175	4	US-08-937-155-7
4	892	93.3	175	4	US-09-323-998E-7
5	892	93.3	175	4	US-08-624-125-6
6	892	93.3	175	4	US-08-937-155-6
7	892	93.3	175	4	US-09-323-998E-6
8	892	93.3	175	1	US-08-096-623A-18
9	892	93.3	175	1	US-08-624-125-5
10	892	93.3	175	2	US-08-663-310-11
11	892	93.3	175	2	US-09-006-491-11
12	892	93.3	175	3	US-09-335-919-11
13	892	93.3	175	4	US-08-937-155-5
14	892	93.3	175	4	US-09-323-998E-5
15	892	93.3	175	2	US-08-663-310-4
16	892	93.3	175	2	US-09-006-491-4
17	892	93.3	175	3	US-09-335-919-4
18	892	93.3	175	3	US-08-980-832-34
19	892	93.3	175	4	US-09-920-923B-34
20	892	93.3	175	1	US-08-624-125-8
21	892	93.3	175	4	US-08-937-155-8
22	892	93.3	175	4	US-09-323-998E-8
23	892	93.3	175	3	US-08-660-645A-9
24	892	93.3	175	3	US-09-298-718-9
25	892	93.3	175	3	US-09-546-969-9
26	892	93.3	175	3	US-08-980-832-6
27	892	93.3	175	4	US-09-547-267-9

Sequence 6, Appli  
Sequence 14, Appli  
Sequence 2, Appli  
Sequence 15, Appli  
Sequence 16, Appli  
Sequence 4, Appli  
Sequence 10, Appli  
Sequence 13, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 12, Appli  
Sequence 8, Appli  
Sequence 2, Appli  
Sequence 6, Appli  
Sequence 8, Appli  
Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-07-783-705A-6

; Sequence 6, Application US/07783705A  
; Patent No. 5429939  
; GENERAL INFORMATION:  
; APPLICANT: Misawa, No. 5429939ihiko  
; APPLICANT: Kobayashi, Kazuo  
; APPLICANT: Nakamura, Katsumi  
; APPLICANT: Yamano, Shigeo  
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE  
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Ladas & Parry  
; STREET: 26 West 61 Street  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10023

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: N/A

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/783,705A

FILING DATE: 19911023

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 1-103078

FILING DATE: 21-APR-1989

APPLICATION NUMBER: JP 2-53225

FILING DATE: 05-MAR-1990

APPLICATION NUMBER: US 07/519,011

FILING DATE: 19-APR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Schwadron, Janet I.

REGISTRATION NUMBER: 33,778

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-708-1935

TELEFAX: 212-246-5959

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 175 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-783-705A-6

Query Match 93.3%; Score 892; DB 1; Length 175;

Best Local Similarity 91.4%; Pred. No. 1.1e-101;  
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Y 1 MEMIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAI 60  
b 1 MEMIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAA 60  
Y 61 VSIALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDLGHVHQRWPPFYIPRKGYLKELYMAH 120  
b 61 LSILLIYLGSTGWPLOWIGAGMTAYGLLYFMVHDLGHVHQRWPPFYIPRKGYLKELYMAH 120  
Y 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175  
b 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175

RESULT 2  
S-08-624-125-7  
Sequence 7, Application US/08624125  
Patent No. 5744341  
GENERAL INFORMATION:  
APPLICANT: CUNNINGHAM JR., FRANCIS X.  
APPLICANT: SUN, ZAIREN  
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND  
TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/624,125  
FILING DATE: 29-MAR-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: KELBER, STEVEN B.  
REGISTRATION NUMBER: 30,073  
REFERENCE/DOCKET NUMBER: 2747-063-27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 7:  
APPLICATION NUMBER: US/08/624,125  
FILING DATE: 29-MAR-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: KELBER, STEVEN B.  
REGISTRATION NUMBER: 30,073  
REFERENCE/DOCKET NUMBER: 2747-063-27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 175 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
S-08-624-125-7

Query Match 93.3%; Score 892; DB 1; Length 175;  
Best Local Similarity 91.4%; Pred. No. 1.1e-101;  
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Y 1 MEMIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAI 60  
b 1 MEMIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAA 60  
Y 61 VSIALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDLGHVHQRWPPFYIPRKGYLKELYMAH 120  
b 61 LSILLIYLGSTGWPLOWIGAGMTAYGLLYFMVHDLGHVHQRWPPFYIPRKGYLKELYMAH 120  
Y 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175  
b 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175

Db 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175

RESULT 3  
US-08-937-155-7  
Sequence 7, Application US/08937155  
Patent No. 6524811  
GENERAL INFORMATION:  
APPLICANT: CUNNINGHAM JR., FRANCIS X.  
APPLICANT: SUN, ZAIREN  
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND  
TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/937,155  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/624,125  
FILING DATE: 29-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: KELBER, STEVEN B.  
REGISTRATION NUMBER: 30,073  
REFERENCE/DOCKET NUMBER: 2747-063-27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 175 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-937-155-7

Query Match 93.3%; Score 892; DB 4; Length 175;  
Best Local Similarity 91.4%; Pred. No. 1.1e-101;  
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

Qy 1 MEMIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAI 60  
Db 1 MEMIWNALIVFTVVGMEVVAALAHKYIMHGKMGWHLSSHEPRKGAPEVNDLYAVVFAA 60  
Qy 61 VSIALYFGSTGIWPLQWIGAGMTAYGLLYFMVHDLGHVHQRWPPFYIPRKGYLKELYMAH 120  
Db 61 LSILLIYLGSTGWPLOWIGAGMTAYGLLYFMVHDLGHVHQRWPPFYIPRKGYLKELYMAH 120  
Qy 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175  
Db 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRERHGAARDAQGGDEPASGK 175

RESULT 4  
US-09-323-998E-7  
Sequence 7, Application US/09323998E  
Patent No. 6642021  
GENERAL INFORMATION:  
APPLICANT: CUNNINGHAM JR., FRANCIS X.  
APPLICANT: SUN, ZAIREN

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND  
TITLE OF INVENTION: METHODS OF USE THEREOF  
FILE REFERENCE: 108172-09019  
CURRENT APPLICATION NUMBER: US/09/323,998E  
CURRENT FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: 09/088,724  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 09/088,725  
PRIOR FILING DATE: 1998-06-02  
NUMBER OF SEQ ID NOS: 62  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 175  
TYPE: PRT  
ORGANISM: Erwinia uredovora  
-09-323-998B-7

Query Match 93.3%; Score 892; DB 4; Length 175;  
Best Local Similarity 91.4%; Pred. No. 1.1e-101; Indels 0; Gaps 0;  
Matches 160; Conservative 6; Mismatches 9;

1 MLMTWNAIIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVPAL 60  
|||||  
1 MLMTWNAIIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVPAA 60  
|||||

61 VSIALLYFGSTGIPWLOIGAGTAYGLLYFMVHDGLVHQWPPRYIPRKGYLKRLYMAH 120  
|||||

61 LSILLIYLGSTGMPLOIGAGTAYGLLYFMVHDGLVHQWPPRYIPRKGYLKRLYMAH 120  
|||||

121 RMHVAVRKGGCVSFGFLYAPPLSKQATLRERHAARSAGARDEQGVDTSSGK 175  
|||||

121 RMHVAVRKGGCVSFGFLYAPPLSKQATLRERHAGARDAQGGEDBPASGK 175  
|||||

SULT 5  
-08-624-125-6  
Sequence 6, Application US/08624125  
Patent No. 5744341  
GENERAL INFORMATION:  
APPLICANT: CUNNINGHAM JR., FRANCIS X.  
APPLICANT: SUN, ZAIREN  
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND  
TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/624,125  
FILING DATE: 29-MAR-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: KELBER, STEVEN B.  
REGISTRATION NUMBER: 30,073  
REFERENCE/DOCKET NUMBER: 2747-063-27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 175 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-624-125-6

Query Match 68.4%; Score 654; DB 1; Length 175;  
Best Local Similarity 68.2%; Pred. No. 2e-72;  
Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;

6 NALIVFTVTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVPAL 65  
|||||

3 NSLIVLSVIAEGIAAATTHRYIMHGNGWGHLSHHEPRKGVFELNDLPVAVFAGVAIAL 62  
|||||

66 IYFGSTGIPWLOIGAGTAYGLLYFMVHDGLVHQWPPRYIPRKGYLKRLYMAH 125  
|||||

63 IAVGTAGVPLQWIGCGTYGLLYFLVHDGLVHQWPPRYIPRKGYLKRLYMAH 122  
|||||

126 VRKGGCVSFGFLYAPPLSKQATLRERHA--ARSGAARDEQGVDTSSS 173  
|||||

123 VRKGGCVSFGFLYAPPLSKQATLRERHGRPPKRDAAKDRPDAAASPS 172  
|||||

RESULT 6  
US-08-937-155-6  
Sequence 6, Application US/08937155  
Patent No. 6524811  
GENERAL INFORMATION:  
APPLICANT: CUNNINGHAM JR., FRANCIS X.  
APPLICANT: SUN, ZAIREN  
TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND  
TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/937,155  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/624,125  
FILING DATE: 29-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: KELBER, STEVEN B.  
REGISTRATION NUMBER: 30,073  
REFERENCE/DOCKET NUMBER: 2747-063-27  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 175 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-937-155-6

Query Match 68.4%; Score 654; DB 4; Length 175;  
Best Local Similarity 68.2%; Pred. No. 2e-72;  
Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;

6 NALIVFTVTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVPAL 65  
|||||



3 NSLIVLTVIAMEGIAAFTTHRYIMEGWGRWHESHHTPRKGVFLNDLFAVVFAGVIAL 62  
66 IYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFPRYIPRGYKRLYMAHRMHA 125  
63 IAVGTAGVWPLQWIGCMVTYGLLYFLVHDGLVHORWPFPRYIPRGYKRLYVAHRLHA 122  
126 VRGKGCVSFGFLYAPPLSKLQATLRERHA--ARSGAARDEODGVDTS 173  
123 VRGKGCVSFGFIYARKPADLQAILRERHGRPPKRDAAKRDPAASPSS 172

## RESULT 7

S-09-323-998E-6

Sequence 6, Application US/09323998E

Patent No. 6642021

GENERAL INFORMATION:

APPLICANT: CUNNINGHAM JR., FRANCIS X.

APPLICANT: SUN, ZAIEN

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND METABOLISM AND

METHODS OF USE THEREOF

FILE REFERENCE: 108172-09019

CURRENT APPLICATION NUMBER: US/09/323,998E

PRIOR FILING DATE: 1999-06-02

PRIOR APPLICATION NUMBER: 09/088,724

PRIOR FILING DATE: 1998-06-02

PRIOR APPLICATION NUMBER: 09/088,725

PRIOR FILING DATE: 1998-06-02

NUMBER OF SEQ ID NOS: 62

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 6

LENGTH: 175

TYPE: PRT

ORGANISM: Erwinia herbicola

S-09-323-998E-6

Query Match 68.4%; Score 654; DB 4; Length 175;  
Best Local Similarity 68.2%; Pred. No. 2e-72;  
Matches 116; Conservative 21; Mismatches 31; Indels 2; Gaps 1;

Y 6 NALIVFTVVGMEVVAALAHKYIMHGWGWHLSHHEPRKGAFAFVNDLYAVVFATVIAL 65  
b 3 NSLIVLTVIAMEGIAAFTTHRYIMEGWGRWHESHHTPRKGVFLNDLFAVVFAGVIAL 62  
Y 66 IYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFPRYIPRGYKRLYMAHRMHA 125  
b 63 IAVGTAGVWPLQWIGCMVTYGLLYFLVHDGLVHORWPFPRYIPRGYKRLYVAHRLHA 122  
Y 126 VRGKGCVSFGFLYAPPLSKLQATLRERHA--ARSGAARDEODGVDTS 173  
b 123 VRGKGCVSFGFIYARKPADLQAILRERHGRPPKRDAAKRDPAASPSS 172

## RESULT 8

S-08-096-623A-18

Sequence 18, Application US/08096623A

Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.

APPLICANT: Brinkhaus, Friedhelm L.

APPLICANT: Mukharji, Indrani

APPLICANT: Proffitt, John H.

APPLICANT: Yarger, James G.

APPLICANT: Yen, Huei-Che B.

TITLE OF INVENTION: Biosynthesis of Zeaxanthin and

Glycosylated Zeaxanthin in Genetically Engineered Hosts

NUMBER OF SEQUENCES: 104

CORRESPONDENCE ADDRESS:

ADDRESSEE: Welsh &amp; Katz, Ltd

STREET: 120 S. Riverside Plaza, 22nd Floor

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,623A  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,061  
FILING DATE: 09-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/662,921  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/562,674  
FILING DATE: 03-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/525,551  
FILING DATE: 18-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,613  
FILING DATE: 02-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Ganson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER: AMO-006.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 655-1500  
TELEFAX: (312) 655-1501  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 176 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-096-623A-18

Query Match 68.3%; Score 653; DB 1; Length 176;

Best Local Similarity 68.2%; Pred. No. 2.7e-72;

Matches 116; Conservative 20; Mismatches 32; Indels 2; Gaps 1;

Qy 6 NALIVFTVVGMEVVAALAHKYIMHGWGWHLSHHEPRKGAFAFVNDLYAVVFATVIAL 65  
Db 4 NSLIVLTVIAMEGIAAFTTHRYIMEGWGRWHEPHHTPRKGVFLNDLFAVVFAGVIAL 63  
Qy 66 IYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFPRYIPRGYKRLYMAHRMHA 125  
Db 64 IAVGTAGVWPLQWIGCMVTYGLLYFLVHDGLVHORWPFPRYIPRGYKRLYVAHRLHA 123  
Qy 126 VRGKGCVSFGFLYAPPLSKLQATLRERHA--ARSGAARDEODGVDTS 173  
Db 124 VRGKGCVSFGFIYARKPADLQAILRERHGRPPKRDAAKRDPAASPSS 173

## RESULT 9

US-08-624-125-5

Sequence 5, Application US/08624125

Patent No. 5744341

GENERAL INFORMATION:

APPLICANT: CUNNINGHAM JR., FRANCIS X.

APPLICANT: SUN, ZAIEN

TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND

METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: P.C.

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &amp; NEUSTADT,

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

## REFERENCES

APPLICATION NUMBER: US 08/663,310

ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/335,919  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/663,310

RESULT 13  
 US-08-937-155-5  
 Sequence 5, Application US/08937155  
 Patent No. 6524811  
 GENERAL INFORMATION:  
 APPLICANT: CUNNINGHAM JR., FRANCIS X.  
 APPLICANT: SUN ZAIREN  
 TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND  
 TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ADDRESSEE: P.C.  
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
 CITY: ARLINGTON  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/937,155  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/624,125  
 FILING DATE: 29-MAR-1996  
 ATTORNEY/AGENT INFORMATION:



---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:26:38 ; Search time 20.6908 Seconds  
(without alignments)  
2389.754 Million cell updates/sec

file: US-09-941-947a-36  
irect score: 956  
quence: 1 MLINWALIVFTVGVGEV.....ARSGAARDQGDVTSSGK 175

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

tal number of hits satisfying chosen parameters: 1586107

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : A Geneseq 29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	956	100.0	175	5	Aae22316 Pantoea s
2	956	100.0	175	6	Aao16023 Pantoea s
3	956	100.0	175	6	Abp96690 Pantoea s
4	892	93.3	175	2	Aar07468 Polypepti
5	892	93.3	175	2	Aaw87892 Protein e
6	860	90.0	175	5	Abg93890 Erwinia u
7	654	68.4	176	2	Aar13988 Beta-car
8	642	67.2	175	5	Abg93889 Erwinia n
9	499	52.2	162	2	Aar79061 3 hydroxy
10	499	52.2	162	2	Aaw87886 Protein e
11	495	51.8	162	2	AAR79059 3 hydroxy
12	495	51.8	162	2	Aaw82259 C. utilis
13	495	51.8	162	2	Aaw69536 crtE2396
14	495	51.8	162	2	Aaw87883 Protein e
15	495	51.8	219	2	Aaw98197 SSU/beta-
16	495	51.8	219	5	Aag78480 SSU/crt2
17	489	51.2	162	5	Abg93888 Alkaligen
18	487	50.9	169	2	Aaw06518 Flavobact
19	487	50.9	169	6	Aaw69534 Flavobact
20	487	50.9	169	6	Abu97245 Enzyme po
21	478	50.0	162	5	Abg93891 Agrobacte
22	271	28.3	303	5	ABE93795 Herbicida
23	270	28.2	322	3	Aab11111 H. pluvia
24	268.5	28.1	255	3	Aay32322 Soybean b
25	268.5	28.1	314	3	Aay32321 Soybean b

## ALIGNMENTS

## RESULT 1

Aae22316

ID AAE22316 standard; protein; 175 AA.

XX

AC AAE22316;

XX

DT 25-JUL-2002 (first entry)

XX

DE Pantoea stewartii beta-carotene hydrolase (CrtZ) enzyme.

XX

KW Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;

KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;

KW aquaculture; enzyme; beta-carotene hydrolase; CrtZ.

XX

OS Pantoea stewartii.

XX

PN WO200218617-A2.

XX

PD 07-MAR-2002.

XX

PF 04-SEP-2001; 2001WO-US027420.

XX

PR 01-SEP-2000; 2000US-0229858P.

XX

PR 01-SEP-2000; 2000US-0229907P.

XX

PA (DUFO) DU PONT DE NEMOURS & CO E I.

XX

PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX

PI Odom JM, Picataggio SK, Rouviere PE;

XX

XX WPI: 2002-351711/38.

XX

DR N-PSDB; AAD35514.

XX

DR Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by

XX using microorganisms having a nucleic acid molecule encoding enzymes in

XX the carotenoid biosynthetic pathway and which metabolize single carbon

XX substrates.

XX

PS Claim 23; Page 146; 156pp; English.

XX

XX The invention relates to a method for producing carotenoid compounds. The

CC method comprises a transformed metabolizing host cell, comprising

CC suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule

CC encoding an enzyme in the carotenoid biosynthetic pathway, under the

CC control of regulatory sequences, and contacting the host cell with carbon

CC substrate to produce a carotenoid compound. The method is useful for

CC producing carotenoid compounds such as antheraxanthin and astaxanthin, by

CC using microorganism having a nucleic acid molecule encoding enzymes in

CC

C the carotenoid biosynthetic pathway and which metabolise single carbon  
C substrates. The carotenoids have potent anti-oxidant properties useful in  
C diet, and aquaculture elements. The carotenoids are also useful as  
C intermediates in the synthesis of steroids flavours and fragrances and  
C compounds for potential electro-optic applications. The present sequence  
C is Pantoea stewartii beta-carotene hydrolase (CrtZ) enzyme used in the  
C invention

X Sequence 175 AA;

Query Match 100.0%; Score 956; DB 5; Length 175;  
Best Local Similarity 100.0%; Pred. No. 9.7e-108;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVFAI 60  
b 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVFAI 60  
Y 61 VSTALIYFGSTGIWPLQWIGAGTAYGLLYFMVHDGLVHQRPFFRYIPRKGYLKELYMAH 120  
b 61 VSTALIYFGSTGIWPLQWIGAGTAYGLLYFMVHDGLVHQRPFFRYIPRKGYLKELYMAH 120  
Y 121 RMEHAVRGEGCVSFGFLYAPPLSKLQATLRERHAARSGAARDEODGVDTSSSGK 175  
b 121 RMEHAVRGEGCVSFGFLYAPPLSKLQATLRERHAARSGAARDEODGVDTSSSGK 175

## RESULT 2

AO16023  
D AAO16023 standard; protein; 175 AA.

C AAO16023;

T 20-FEB-2003 (first entry)

E Pantoea stewartii Beta-carotene hydroxylase.

X Carotenoid; crt.

X Pantoea stewartii.

X WO200279395-A2.

X 10-OCT-2002.

X 25-JAN-2002; 2002WO-US0002124.

X 26-JAN-2001; 2001US-0264329P.

X 04-MAY-2001; 2001US-0288984P.

X (CRGI ) CARGILL INC.

X De Souza ML, Kollmann SR, May CA, Schroeder WA;

X WPI; 2003-075455/07.

X N-PSDB; AET14195.

X Novel isolated nucleic acid useful e.g. to engineer host cells with the  
X ability to produce particular carotenoids and polypeptides useful in cell  
X -free systems to make particular carotenoids.

X Claim 38; Page 65-66; 74pp; English.

X The invention comprises the amino acid and coding sequence of a number of  
X carotenoid (crt)-related proteins. The crt-related DNA and protein  
X sequences of the invention are useful for engineering cells which are  
X able to produce carotenoids. The present amino acid sequence represents a  
X crt-related protein of the invention

X Sequence 175 AA;

Query Match 100.0%; Score 956; DB 6; Length 175;

Best Local Similarity 100.0%; Pred. No. 9.7e-108;

Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVFAI 60  
Db 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVFAI 60  
Qy 61 VSTALIYFGSTGIWPLQWIGAGTAYGLLYFMVHDGLVHQRPFFRYIPRKGYLKELYMAH 120  
Db 61 VSTALIYFGSTGIWPLQWIGAGTAYGLLYFMVHDGLVHQRPFFRYIPRKGYLKELYMAH 120

Qy 121 RMEHAVRGEGCVSFGFLYAPPLSKLQATLRERHAARSGAARDEODGVDTSSSGK 175

Db 121 RMEHAVRGEGCVSFGFLYAPPLSKLQATLRERHAARSGAARDEODGVDTSSSGK 175

## RESULT 3

ABP96690

ID ABP96690 standard; protein; 175 AA.

XX ABP96690;

DT 03-JUN-2003 (first entry)

XX Pantoea stewartii beta-carotene hydroxylase SEQ ID NO:12.

XX Pantoea stewartii; carotenoid biosynthetic enzyme; crtB; crtX; crtY;  
XX crtI; crtZ; beta-carotene hydroxylase; enzyme; phytoene;  
XX carotenoid.

XX Pantoea stewartii.

XX WO2003016503-A2.

XX 27-FEB-2003.

XX 15-AUG-2002; 2002WO-US026647.

XX 15-AUG-2001; 2001US-0312646P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PE;

XX WPI; 2003-268323/26.

XX N-PSDB; ACC44764.

XX Novel nucleic acid molecule isolated from Pantoea stewartii encoding a  
XX carotenoid biosynthetic enzyme, useful for regulating carotenoid  
XX biosynthesis in an organism.

XX Claim 4; Page 67-68; 68pp; English.

XX The present invention describes Pantoea stewartii carotenoid biosynthetic  
XX enzymes (I). More specifically described are the geranylgeranyl  
XX pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX),  
XX lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase  
XX (crtB) and beta-carotene hydroxylase (crtZ) enzymes (see ABP96685 to  
XX ABP96690) encoded by ACC44759 to ACC44764. (I) can be used for regulating  
XX carotenoid biosynthesis in an organism, by over-expressing (I) in an  
XX organism, such that the carotenoid biosynthesis is altered in the  
XX organism. (I) and the genes encoding (I) are useful for converting  
XX phytoene to the carotenoids, for creating recombinant organisms that have  
XX the ability to produce various carotenoid compounds, and also for  
XX enhancing or manipulating carotenoid compounds. (I) can also be used for  
XX producing gene products having enhanced or altered activity

XX Sequence 175 AA;

Query Match 100.0%; Score 956; DB 6; Length 175;

Best Local Similarity 100.0%; Pred. No. 9.7e-108;  
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLMIWNALIVFTVVGMEVVAALAHKYIMHGNGWGHLSHHEPRKGAFEVNDLYAVVFAI 60

1 MLTWNALIVFTVVGMEVVAALAHKYIMHGWMGWLHSHHEPRKGAPEVNDLYAVVFAI 60  
61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120  
61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120  
121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRRHHAARSGAARDQGDGVDTSSTSGK 175  
121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRRHHAARSGAARDQGDGVDTSSTSGK 175

RESULT 4  
LR07468  
2 AAR07468 standard; protein; 175 AA.  
AAR07468;  
24-OCT-2003 (revised)  
28-JAN-1991 (first entry)  
Polypeptide with enzymatic activity for the conversion of beta-carotene  
into zeaxanthin.  
Carotenoid biosynthesis; vitamin A; cancer; food coloring.  
Pantoea ananatis.  
EP393690-A.  
24-OCT-1990.  
20-APR-1990; 90SP-00107493.  
21-APR-1989; 83JP-00103078.  
05-MAR-1990; 90JP-00053255.  
(KIRI ) KIRIN BEER KK.  
Misawa N, Kobayashi K, Nakamura K;  
WPI; 1990-322212/43.  
N-PSDB; AAQ06298.  
DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn.  
of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.  
Claim 6; Fig 6; 40pp; English.  
Gene products are useful for the synthesis of carotenoids, useful as food  
coloring, vitamin A precursor, and possibly in prevention of cancer. See  
also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)

Sequence 175 AA;  
Query Match 93.3%; Score 892; DB 2; Length 175;  
Best Local Similarity 91.4%; Pred. No. 6e-100;  
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
1 MLTWNALIVFTVVGMEVVAALAHKYIMHGWMGWLHSHHEPRKGAPEVNDLYAVVFAI 60  
1 MLTWNALIVFTVVGMEVVAALAHKYIMHGWMGWLHSHHEPRKGAPEVNDLYAVVFAA 60  
61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120  
61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120  
121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRRHHAARSGAARDQGDGVDTSSTSGK 175  
121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRRHHAARSGAARDQGDGVDTSSTSGK 175

AAW87892  
ID AAW87892 standard; protein; 175 AA.  
XX  
AC AAW87892;  
XX  
DT 17-OCT-2003 (revised)  
DT 10-MAR-1999 (first entry)  
XX  
XX Protein encoded by the carotenoid biosynthesis gene crtZ.  
XX  
XX Carotenoid biosynthesis; astaxanthin diglucoside; crtE gene; crtB gene;  
XX crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside;  
XX carotenoid glycoside; crtY gene; crtZ gene; crtX gene; crtW gene;  
XX food additive.  
XX  
XX Pantoea ananatis.  
XX  
XX JP10327865-A.  
XX  
PD 15-DEC-1998.  
XX  
XX 29-MAY-1997; 97JP-00140460.  
XX  
XX 29-MAY-1997; 97JP-00140460.  
XX  
XX (KIRI ) KIRIN BREWERY KK.  
XX (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.  
XX WPI; 1999-099030/09.  
XX N-PSDB; AAV84084.  
XX  
XX New carotenoid glucoside(s) - used as food additives.  
XX  
XX Disclosure; Page 23-24; 26pp; Japanese.  
XX  
XX The present sequence represents a protein involved in carotenoid  
XX biosynthesis. The specification describes astaxanthin diglucosides and  
XX adonixanthin-3'-glucosides. The specification also describes a method for  
XX the preparation of a carotenoid glycoside, in which all, or part of,  
XX carotenoid biosynthesis genes crtE, crtB, crtI, crtY, crtZ, crtX or crtW  
XX are introduced to a microbe or plant and expressed. The transformed  
XX organism is cultured and astaxanthin diglucosides, adonixanthin-3'-  
XX glucosides, and/or astaxanthin monoglucoside are collected. The  
XX carotenoid glucosides are used as food additives. (Updated on 17-OCT-2003  
XX to standardise OS field)  
XX  
SQ Sequence 175 AA;  
Query Match 93.3%; Score 892; DB 2; Length 175;  
Best Local Similarity 91.4%; Pred. No. 6e-100;  
Matches 160; Conservative 6; Mismatches 9; Indels 0; Gaps 0;  
QY 1 MLTWNALIVFTVVGMEVVAALAHKYIMHGWMGWLHSHHEPRKGAPEVNDLYAVVFAI 60  
DB 1 MLTWNALIVFTVVGMEVVAALAHKYIMHGWMGWLHSHHEPRKGAPEVNDLYAVVFAA 60  
QY 61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120  
DB 61 VSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPFYIPRKGYLKRLYMAH 120  
QY 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRRHHAARSGAARDQGDGVDTSSTSGK 175  
DB 121 RMHHAVRGKGCVSFGFLYAPPLSKLOATLRRHHAARSGAARDQGDGVDTSSTSGK 175

RESULT 6  
ABG93890  
ID ABG93890 standard; protein; 175 AA.  
XX  
AC ABG93890;  
XX  
XX 29-AUG-2003 (revised)  
DT 28-NOV-2002 (first entry)



X Erwinia uredovora beta-carotene hydroxylase.  
 X Epsilon-cyclase; carotenoid; isopentenyl pyrophosphate isomerase;  
 W beta-carotene hydroxylase; plant; enzyme.  
 X Pantoea agglomerans.  
 S US2002102631-A1.  
 N 01-AUG-2002.  
 D 02-JUN-1999; 99US-00323998.  
 F 29-MAR-1996; 96US-00624125.  
 R 25-SEP-1997; 97US-00937155.  
 R 02-JUN-1998; 98US-00088724.  
 R 02-JUN-1998; 98US-00088725.  
 X (CUNN/) CUNNINGHAM F X.  
 A (SUNZ/) SUN Z.  
 X Cunningham FX, Sun Z;  
 X WPI; 2002-690609/74.  
 X Producing, or enhancing production of a carotenoid in a host cell by  
 T transforming the cell with a vector encoding a epsilon cyclase enzyme.  
 X Example 1; Fig 6; 85pp; English.  
 X The invention relates to a method of producing or enhancing production of  
 X a carotenoid in a host cell. The method comprises inserting into the cell  
 X a vector comprising a heterologous nucleic acid sequence encoding a  
 X protein with epsilon-cyclase enzyme activity operably linked to a  
 X promoter, and expressing the heterologous sequence to produce the  
 X protein. ABG93886-ABG93914 represent epsilon-cyclase, isopentenyl  
 X pyrophosphate isomerase and beta-carotene hydroxylase proteins of the  
 X invention. (Updated on 29-AUG-2003 to standardise OS field)  
 X Sequence 175 AA;  
 SQ Query Match 90.0%; Score 860; DB 5; Length 175;  
 Best Local Similarity 89.1%; Pred. No. 4.7e-96;  
 Matches 156; Conservative 6; Mismatches 13; Indels 0; Gaps 0;  
 QY 1 MLWLNALLIVFTVVGMEVVAALAHKYIMHGNGWHLSHHEPRKGAPEVNDLYAVVPAI 60  
 DB 1 MLWLNALLIVFTV:ZMEVIAALAHKYIMHGNGWHLSHHEPRKGAPEVNDLYAVVPA 60  
 QY 61 VSLIIFYGSGTGMPLQWIGAGMTAYGLLYFMVHDLVHQRWPPRYIPRKGYLKRLYMAH 120  
 DB 61 LSLIIFYGSGTGMPLQWIGAGMTAYGLLYFMVHDLVHQRWPPRYIPRKGYLKRLYMAH 120  
 QY 121 RMHVAVRZEGCVSGFGFLYAPPLSKLQATLREHRAARSGAARDQDQGVDTSSGK 175  
 DB 121 RMHVAVRZEGCVSGFGFLYAPPLSKLQATLREHRAARSGAARDQDQGVDTSSGK 175  
 RESULT 7  
 ID AAR13988 standard; protein; 176 AA.  
 XX AAR13988;  
 AC AAR13988;  
 XX 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 26-NOV-1991 (first entry)  
 XX Beta-carotene hydroxylase - variant.  
 XX GGPP; carotenoid; phytoene; zeaxanthin; lycopene; ss.  
 XX

OS Pantoea agglomerans; EHO-10 (E. vulneris - ATCC 39368).  
 XX W09113078-A.  
 XX 05-SEP-1991.  
 XX 04-MAR-1991; 91WO-US001458.  
 XX 02-MAR-1990; 90US-00487613.  
 PR 18-MAY-1990; 90US-00525551.  
 PR 03-AUG-1990; 90US-00562674.  
 PR 28-FEB-1991; 91US-00662921.  
 XX {STAD } AMOCO CORP.  
 PA Ausich RL, Brinkhaus PL, Mukharji I, Proffitt JH, Yarger JG;  
 PI Yen HC;  
 XX WPI; 1991-281410/38.  
 DR N-PSDB; AAQ13725.  
 XX Biosynthesis of carotenoid(s) in genetically engineered hosts - using DNA  
 PT encoding enzymes from Erwinia herbicola.  
 XX Disclosure; Fig 21(1-3); 313pp; English.  
 XX There are a total of six relevant genes in a 7900 bp region that cause E.  
 CC coli cells to produce GGPP and the carotenoids phytoene through  
 CC zeaxanthin diglucoside, which is the final prod. identified in the  
 CC carotenoid pathway contd. in plasmid pARC376 (contg. a ca. 13 kb  
 CC chromosomal DNA fragment isolated by Perry et al., J. Bacteriol., 168:607  
 CC (1986)). The genes for geranylgeranyl pyrophosphate (GGPP) synthase,  
 CC phytoene synthase, phytoene dehydrogenase-4H, lycopene cyclase, beta-  
 CC carotene hydroxylase, and zeaxanthin glycosylase are represented in  
 CC AAQ13716, AAQ13718, AAQ13719, AAQ13722, AAQ13724 and AAQ13726  
 CC respectively. The native sequence (AAQ13724) was genetically engineered.  
 CC At the 5' end of the gene, codons encoding the native second and third  
 CC amino acid have been changed from Leu-Val to Val-Leu. Recombinant  
 CC expression plasmids can be used to produce large amts. of the enzymes and  
 CC hence large amts. of the carotenoids which they synthesise. (Updated on  
 CC 25-MAR-2003 to correct PF field.) (Updated on 25-MAR-2003 to correct PR  
 CC field.) (Updated on 24-OCT-2003 to standardise OS field)  
 XX Sequence 176 AA;  
 SQ Query Match 68.4%; Score 654; DB 2; Length 176;  
 Best Local Similarity 67.4%; Pred. No. 5.5e-71;  
 Matches 116; Conservative 21; Mismatches 33; Indels 2; Gaps 1;  
 QY 4 IWNALIVFTVVGMEVVAALAHKYIMHGNGWHLSHHEPRKGAPEVNDLYAVVPAI 63  
 DB 2 VLNSLIIVLTIVAMEGIAAFTTHYIMHGNGWHLSHHEPRKGVFELNDLFVAVVPAI 61  
 QY 64 ALIVFGSTGIWPLQWIGAGMTAYGLLYFMVHDLVHQRWPPRYIPRKGYLKRLYMAH 123  
 DB 62 ALIATGATAGVWPLQWIGAGMTAYGLLYFMVHDLVHQRWPPRYIPRKGYLKRLYMAH 121  
 QY 124 HAVRZEGCVSGFGFLYAPPLSKLQATLREHRA--ARSGAARDQDQGVDTSSS 173  
 DB 122 HAVRZEGCVSGFGFIYARKPADLQAILREHRCRPPKRDAAKORPDASPS 173  
 RESULT 8  
 ID ABG93889 standard; protein; 175 AA.  
 XX ABG93889  
 AC ABG93889;  
 XX 29-AUG-2003 (revised)  
 DT 28-NOV-2002 (first entry)  
 XX Erwinia herbicola beta-carotene hydroxylase.  
 XX

W Epsilon-cyclase; carotenoid; isopentenyl pyrophosphate isomerase;  
 W beta-carotene hydroxylase; plant; enzyme.

S Pantoea agglomerans.

X US2002102631-A1.

X 01-AUG-2002.

X 02-JUN-1999; 99US-00323998.

X 29-MAR-1996; 96US-00624125.

X 25-SEP-1997; 97US-00937155.

X 02-JUN-1998; 98US-00088724.

X 02-JUN-1998; 98US-00088725.

X (CUNN/) CUNNINGHAM F X.

X (SUNZ/) SUN Z.

X Cunningham FX, Sun Z;

X WPI; 2002-590609/74.

X Producing, or enhancing production of a carotenoid in a host cell by

X transforming the cell with a vector encoding a epsilon cyclase enzyme.

X Example 1; Fig 6; 85pp; English.

X The invention relates to a method of producing or enhancing production of  
 X a carotenoid in a host cell. The method comprises inserting into the cell  
 X a vector comprising a heterologous nucleic acid sequence encoding a  
 X protein with epsilon-cyclase enzyme activity operably linked to a  
 X promoter, and expressing the heterologous sequence to produce the  
 X protein. ABG93866-ABG93914 represent epsilon-cyclase, isopentenyl  
 X pyrophosphate isomerase and beta-carotene hydroxylase proteins of the  
 X invention. (Updated on 29-AUG-2003 to standardise OS field)

X Sequence 175 AA;

X Query Match 67.2%; Score 642; DB 5; Length 175;

X Best Local Similarity 66.5%; Pred. No. 1.6e-69;

X Matches 113; Conservative 23; Mismatches 32; Indels 2; Gaps 1;

X Y 6 NALIVFTVVGMEVVAALAHKYIMHG-WGKGNHLSHHEPRKGAPEVNDLYAVVFAIVSTAL 65

X b 3 NSLIVLSVIAEZIAAFTHRYIMHGWNWHESHHTPRKGVFEEDLFAVVFAGVIAL 62

X Y 66 IYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRMHA 125

X b 63 IAVGTAGVWPLQWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRMHA 122

X Y 126 VRGKGCVSFGFLYAPPLSKLQATLBERHA--ARSGAARDEQDVTSSS 173

X b 123 VRGREZCVSFGFIYARKPADLZAILRERHGRPPKDAKORPDAPSPSS 172

X RESULT 9

X AR79061

X D AAR79061 standard; peptide; 162 AA.

X X AAR79061;

X X 28-FEB-1996 (first entry)

X X 3 hydroxy-beta-ionone ring methylene to keto group converting peptide.

X X Xanthophyll; astaxanthine; methylene; keto group; conversion;

X X 3-hydroxy-beta-ionone ring.

X S Alcaligenes.

X N MO9518220-A1.

X X

PD 06-JUL-1995.

XX 26-DEC-1994; 94WO-JP002220.

XX 27-DEC-1993; 93JP-00348737.

PR 05-SEP-1994; 94JP-00235917.

XX (KIRI ) KIRIN BEER KK.

PA (MARI-) MARINE BIOTECHNOLOGY INST CO LTD.

XX Misawa N, Kondo K, Kajiura S, Yokoyama A;

XX WPI; 1995-246386/32.

DR N-PSDB; AAQ99488.

XX DNA's encoding xanthophyll(s) - esp. asta:xanthin and other

PT xanthophyll(s) using e.g. E. coli.

XX Claim 22; Page 95-96; 131pp; Japanese.

XX AAR79058-R790629 are xanthophyll polypeptides. These polypeptides are

CC capable of converting the 4-methylene group of a 3-hydroxy-beta- ionone

CC ring to a 4-keto group in doing so these peptides also add a hydroxyl

CC group to the 3-position carbon-atom of the 4-keto-beta- ionone ring. The

CC DNA sequences may be used in the production of astaxanthine and other

CC keto gp. contg. xanthophylls, the sequences may also be used to transform

CC certain yeasts and other microorganisms

XX SQ Sequence 162 AA;

X Query Match 52.2%; Score 499; DB 2; Length 162;

X Best Local Similarity 57.8%; Pred. No. 3.6e-52;

X Matches 93; Conservative 22; Mismatches 38; Indels 8; Gaps 4;

X Qy 8 LIVFTVVGMEVVAALAHKYIMHG-WGKGNHLSHHEPRKGAPEVNDLYAVVFAIVSIALI 66

X Db 5 LIVVATVLVMELTAYSVHRWIMHGPIGNGWGHKSHHEEDHAEKNDLYGVVFAVLATILF 64

X Qy 67 YFGSTGIWP-LQWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRMHA 125

X Db 65 TVGAYW-WPVLWIAUGMTYGLIYFIEDGLVHORWPPRYIPRKGYFRLYQAHRLHHA 123

X Qy 126 VRGKGCVSFGFLYAPPLSKLQATLBERHAARSGAARDEQD 166

X Db 124 VEGRDHCVSFGFIYAPVDKLDKDLK-----RSGVLRPQDE 159

XX RESULT 10

XX AAW87886

XX ID AAW87886 standard; protein; 162 AA.

XX AC AAW87886;

XX X 10-MAR-1999 (first entry)

XX Protein encoded by the carotenoid biosynthesis gene crtZ.

XX Carotenoid biosynthesis; astaxanthin diglucoside; crtZ gene; crtB gene;

XX crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside;

XX carotenoid glycoside; crtY gene; crtZ gene; crtX gene; crtW gene;

XX food additive.

XX Alcaligenes sp.

XX JPI0327865-A.

XX 15-DEC-1998.

XX 29-MAY-1997; 97JP-00140460.

XX 29-MAY-1997; 97JP-00140460.

XX (KIRI ) KIRIN BREWERY KK.

A (KALY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.  
 X WPI; 1999-09030/09.  
 R N-PSDB; AAV84078.  
 X T New carotenoid glucoside(s) - used as food additives.  
 X S Disclosure; Page 14-15; 26pp; Japanese.  
 X C The present sequence represents a protein involved in carotenoid biosynthesis. The specification describes astaxanthin diglucosides and adonixanthin-3'-glucosides. The specification also describes a method for the preparation of a carotenoid glycoside, in which all, or part of, carotenoid biosynthesis genes crtB, crtE, crtI, crtX, crtZ, crtY or crtW are introduced to a microbe or plant and expressed. The transformed organism is cultured and astaxanthin diglucosides, adonixanthin-3'-glucosides, and/or astaxanthin monoglucoside are collected. The carotenoid glucosides are used as food additives.  
 X C Sequence 162 AA;  
 Q Query Match 52.2%; Score 499; DB 2; Length 162;  
 Best Local Similarity 57.8%; Pred. No. 3.6e-52;  
 Matches 93; Conservative 22; Mismatches 38; Indels 8; Gaps 4;  
 Y 8 LIAFVTVVGMVVAALAHKYIMHG-WGNGWHLSHHEPRKGAPEVNDLYAVVPAIVSIALI 66  
 b 5 LIAVATLVLMELTAYSVHRWIMHGPLGNGWKSHEHDEHDALEKNDLYGVPAIVATIV 64  
 Y 67 YFGSTGIWP--LWIGAGMTAYGLLYFMVHDGLVHQWPPRYIPRKGYLKRLYMAHRMHA 125  
 b 65 TVGAYW-WPVLWMLWIALGMTVYGLIYFVLHDGLVHQWPPRYIPRKGYARLLYQAHRL 123  
 Y 126 VRGKEGCVSFGFLYAPPLSKLQATLREHRAARSGAARDEOD 166  
 b 124 VEGRDHCVSFGFIYAPPVDKLDK-----RSGVLRPQDE 159  
 RESULT 11  
 AAR79059  
 ID AAR79059 standard; peptide; 162 AA.  
 X AAR79059;  
 X 27-AUG-2003 (revised)  
 DT 28-FEB-1996 (first entry)  
 X X  
 DE 3 hydroxy-beta-ionone ring methylene to keto group converting peptide.  
 X Xanthophyll; astaxanthine; methylene; keto group; conversion;  
 KW 3-hydroxy-beta-ionone ring.  
 X Agrobacterium aurantiacum.  
 OS  
 X W09518220-A1.  
 FN  
 PD 06-JUL-1995.  
 X 26-DEC-1994; 94MO-JP002220.  
 X 27-DEC-1993; 93JP-00348737.  
 PR 05-SEP-1994; 94JP-00235917.  
 X X  
 PA (KIRI) KIRIN BEER KK.  
 PA (MARI-) MARINE BIOTECHNOLOGY INST CO LTD.  
 X Misawa N, Kondo K, Kajiwara S, Yokoyama A;  
 X WPI; 1995-246386/32.  
 DR N-PSDB; AAO39486.  
 X DNA's encoding xanthophyll(s) - esp. asta:xanthin and other xanthophyll(s) using e.g. E. coli.  
 PT

XX Claim 20; Page 79-80; 131pp; Japanese.  
 PS  
 XX AAR79058-R790629 are xanthophyll polypeptides. These polypeptides are capable of converting the 4-methylene group of a 3-hydroxy-beta-ionone ring to a 4-keto group in doing so these peptides also add a hydroxyl group to the 3-position carbon-atom of the 4-keto-beta-ionone ring. The DNA sequences may be used in the production of astaxanthine and other keto gp. contg. xanthophylls, the sequences may also be used to transform certain yeasts and other microorganisms. (Updated on 27-AUG-2003 to correct OS field.)  
 X C Sequence 162 AA;  
 Q Query Match 51.8%; Score 495; DB 2; Length 162;  
 Best Local Similarity 58.0%; Pred. No. 1.1e-51;  
 Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;  
 QY 6 NALIVFTVVGMEVVAALAHKYIMHG-WGNGWHLSHHEPRKGAPEVNDLYAVVPAIVSIA 64  
 b 3 NFLIVATLVLMELTAYSVHRWIMHGPLGNGWKSHEHDEHDALEKNDLYGLVPAIVATV 62  
 QY 65 LIYFGSTGIWP--LWIGAGMTAYGLLYFMVHDGLVHQWPPRYIPRKGYLKRLYMAHRM 122  
 b 63 LFTVG--WIWAPVLWMLWIALGMTVYGLIYFVLHDGLVHQWPPRYIPRKGYARLLYQAHRL 120  
 QY 123 HHAYRGKEGCVSFGFLYAPPLSKLQATLREHRAARSGAARDE 164  
 b 121 HHAVEGRDHCVSFGFIYAPPVDKLDK-----MSGVLRAR 157  
 RESULT 12  
 AAW82259  
 ID AAW82259 standard; protein; 162 AA.  
 X AAW82259;  
 X 17-OCT-2003 (revised)  
 DT 16-JUL-1999 (first entry)  
 X C. utilis crtZ protein.  
 DE HMG-CoA; 3-hydroxy-3-methylglutaryl coenzyme A reductase; crtZ;  
 KW carotenoid.  
 X Pichia jadinii.  
 OS JP10248575-A.  
 FN 22-SEP-1998.  
 X 12-MAR-1997; 97JP-00058012.  
 PR 12-MAR-1997; 97JP-00058012.  
 X (KIRI) KIRIN BREWERY KK.  
 PA WPI; 1998-560727/48.  
 DR N-PSDB; AAV73183.  
 X Gene useful for increase in carotenoid production - and preparation of carotenoid.  
 PT  
 PS Example 2; Fig 18-19; 54pp; Japanese.  
 X This invention describes a novel method for the preparation of carotenoids using genes and proteins isolated from Candida utilis. The invention specifically describes the isolation of a 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase protein. This sequence represents the Candida utilis crtZ protein which is used in the method of the invention. (Updated on 17-OCT-2003 to standardise OS field)  
 X C Sequence 162 AA;  
 QY

Query Match 51.8%; Score 495; DB 2; Length 162;  
 Best Local Similarity 58.0%; Pred. No. 1.1e-51;  
 Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;

6 NALIVFTVVGMEVVAALAHKTYIMHG-WGNGWHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64  
 3 NFLIVVATVLMELTAYSVHRWIMHGPLGWGWHKSHHEHDHLEKNDLYGLVFAVIATV 62

65 LIYFGSTGIWP--LOWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRM 122  
 63 LFTVG--WTAPVLMWIALGMTVYGLIIFVLHDGLVHORWPPRYIPRKGYARRLYQAHRL 120

6 NALIVFTVVGMEVVAALAHKTYIMHG-WGNGWHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64  
 3 NFLIVVATVLMELTAYSVHRWIMHGPLGWGWHKSHHEHDHLEKNDLYGLVFAVIATV 62

65 LIYFGSTGIWP--LOWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRM 122  
 63 LFTVG--WTAPVLMWIALGMTVYGLIIFVLHDGLVHORWPPRYIPRKGYARRLYQAHRL 120

123 HHAVRGKCCVSGFLYAPPLSKLOATLRRHAARSGAARDE 164  
 121 HHAVEGRDHCVSFGFIYAPPPVDKLDK-----MSGVLRAE 157

RESULT 13  
 AAW69536  
 D AAW69536 standard; protein; 162 AA.  
 C AAW69536;  
 K 10-AUG-1999 (first entry)  
 I cr2E2396 gene product beta-carotene hydroxylase.  
 K Carotenoid; pigment; canthaxanthin; R1534; crtB; prephytoene synthase;  
 N crtI; phytoene desaturase; crtY; lycopene cyclase; crtW2396;  
 K beta-carotene beta-oxygenase; food product; fermentation.  
 K Flavobacterium sp.  
 J JP10155497-A.  
 X 16-JUN-1998.  
 D 02-DEC-1997; 97JP-00348653.  
 F 02-DEC-1996; 96EP-00810839.  
 R (HOFF) HOPPMANN LA ROCHE & CO AG F.  
 X WPI; 1998-391048/34.  
 R N-PSDB; AAV40148.  
 X Preparation of carotenoid - comprises fermentation with transformed cell.  
 T Claim 2; Fig 64; 80pp; Japanese.  
 S The invention describes the preparation of carotenoid pigments e.g.  
 C canthaxanthins using a cell transformed by a vector having DNA sequences  
 C (a) to (e) or substantially homologous sequences. (a) a DNA sequence  
 C (crtE) coding GPP synthase of Flavobacterium sp. R1534; (b) a DNA  
 C sequence (crtB) coding prephytoene synthase of Flavobacterium sp. R1534;  
 C (c) a DNA sequence (crtI) coding phytoene desaturase of Flavobacterium  
 C sp. R1534; (d) a DNA sequence (crtY) coding lycopene cyclase of  
 C Flavobacterium sp. R1534; and (e) a DNA sequence (crtW2396) coding beta-  
 C carotene beta-oxygenase of a microbe E-396 (FERM BP-4283). The carotenoid  
 C or a carotenoid mixture can also be used in preparation of food products.  
 C The method is an improved method of fermentation for carotenoid  
 C production  
 X Sequence 162 AA;  
 Q Query Match 51.8%; Score 495; DB 2; Length 162;  
 Best Local Similarity 58.0%; Pred. No. 1.1e-51;  
 Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;

6 NALIVFTVVGMEVVAALAHKTYIMHG-WGNGWHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64  
 3 NFLIVVATVLMELTAYSVHRWIMHGPLGWGWHKSHHEHDHLEKNDLYGLVFAVIATV 62

QY 65 LIYFGSTGIWP--LOWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRM 122  
 DB 63 LFTVG--WTAPVLMWIALGMTVYGLIIFVLHDGLVHORWPPRYIPRKGYARRLYQAHRL 120

QY 123 HHAVRGKCCVSGFLYAPPLSKLOATLRRHAARSGAARDE 164  
 DB 121 HHAVEGRDHCVSFGFIYAPPPVDKLDK-----MSGVLRAE 157

RESULT 14  
 AAW87883  
 ID AAW87883 standard; protein; 162 AA.  
 X AAW87883;  
 AC AAW87883;  
 DT 10-MAR-1999 (first entry)  
 XX Protein encoded by the carotenoid biosynthesis gene crtZ.  
 DE Carotenoid biosynthesis; astaxanthin diglucoside; crtE gene; crtB gene;  
 XX crtI gene; adonixanthin-3'-glucoside; astaxanthin monoglucoside;  
 KW carotenoid glycoside; crtY gene; crtZ gene; crtX gene; crtW gene;  
 XX food additive.  
 OS Agrobacterium aurantiacum.  
 XX JP10327865-A.  
 FN 15-DEC-1998.  
 PD 29-MAY-1997; 97JP-00140460.  
 PF 29-MAY-1997; 97JP-00140460.  
 PR 29-MAY-1997; 97JP-00140460.  
 XX (KIRI) KIRIN BREWERY KK.  
 XX (KAIY-) KAIYO BIOTECHNOLOGY KENKYUSHO KK.  
 PA WPI; 1999-099030/09.  
 DR N-PSDB; AAV84075.  
 XX New carotenoid glucoside(s) - used as food additives.  
 XX Disclosure; Page 11-12; 26pp; Japanese.  
 XX The present sequence represents a protein involved in carotenoid  
 CC biosynthesis. The specification describes astaxanthin diglucosides and  
 CC adonixanthin-3'-glucosides. The specification also describes a method for  
 CC the preparation of a carotenoid glycoside, in which all, or part of,  
 CC carotenoid biosynthesis genes crtE, crtB, crtI, crtY, crtX or crtW  
 CC are introduced to a microbe or plant and expressed. The transformed  
 CC organism is cultured and astaxanthin diglucosides, adonixanthin-3'-  
 CC glucosides, and/or astaxanthin monoglucoside are collected. The  
 CC carotenoid glucosides are used as food additives  
 XX Sequence 162 AA;  
 SQ Query Match 51.8%; Score 495; DB 2; Length 162;  
 Best Local Similarity 58.0%; Pred. No. 1.1e-51;  
 Matches 94; Conservative 19; Mismatches 39; Indels 10; Gaps 4;

6 NALIVFTVVGMEVVAALAHKTYIMHG-WGNGWHLSHHEPRKGAPEVNDLYAVVFAIVSIA 64  
 3 NFLIVVATVLMELTAYSVHRWIMHGPLGWGWHKSHHEHDHLEKNDLYGLVFAVIATV 62

65 LIYFGSTGIWP--LOWIGAGMTAYGLLYFMVHDGLVHORWPPRYIPRKGYLKRLYMAHRM 122  
 DB 63 LFTVG--WTAPVLMWIALGMTVYGLIIFVLHDGLVHORWPPRYIPRKGYARRLYQAHRL 120

123 HHAVRGKCCVSGFLYAPPLSKLOATLRRHAARSGAARDE 164  
 121 HHAVEGRDHCVSFGFIYAPPPVDKLDK-----MSGVLRAE 157



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:33:49 ; Search time 24.3391 Seconds  
(without alignments)  
3837.172 Million cell updates/sec

tie: US-09-941-947a-34

fect score: 1535

quence: 1 MAVGSKSPATASTLFPDKTR.....VTSRMKTYPPRPAHLWQRP1 296

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

abase :

SPTREMBL\_25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriopl:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

sult No.	Score	Query Match	length	DB ID	Description
1	1535	100.0	296	2 Q8GCS0	Q8GCS0 pantoea ste
2	1371	89.3	296	2 Q47846	Q47846 pantoea agg
3	1244	81.0	296	2 Q8VUJ4	Q8VUJ4 pantoea agg
4	804	52.4	311	2 Q93C16	Q93C16 xanthobacte
5	743.5	48.4	335	2 Q9KTX1	Q9KTX1 bradyrhizob
6	686.5	44.7	303	2 P94789	P94789 flavobacter
7	645.5	42.1	304	2 Q9RLH3	Q9RLH3 paracoccus
8	446.5	29.1	346	2 Q8KZ27	Q8KZ27 uncultured
9	403	26.3	379	2 Q8RTY0	Q8RTY0 uncultured
10	395.5	25.8	344	2 Q9JPB3	Q9JPB3 rhodocyclu
11	362	23.6	325	16 Q9RW07	Q9RW07 geinococcus
12	351	22.9	343	2 O50567	O50567 rhodocyclu
13	313.5	20.4	422	10 Q9FV44	Q9FV44 tagetes ere
14	313.5	20.4	436	10 Q9XGW7	Q9XGW7 citrus para
15	310.5	20.2	303	16 Q7U419	Q7U419 synechococ
16	310.5	20.2	436	10 Q9M608	Q9M608 citrus unsh

17	309.5	20.2	399	10 Q8L8H7	Q8L8H7 tagetes ere
18	307.5	20.0	414	10 Q9FSY7	Q9FSY7 helianthus
19	302.5	19.7	414	10 Q9AVV8	Q9AVV8 helianthus
20	301.5	19.6	313	16 Q7V4E9	Q7V4E9 prochloroco
21	300	19.5	336	2 Q50892	Q50892 myxococcus
22	298	19.4	310	16 Q8KCM8	Q8KCM8 chlorobium
23	297.5	19.4	404	10 Q8W014	Q8W014 oryza sativ
24	296.5	19.3	272	10 Q8VWR1	Q8VWR1 oryza sativ
25	293.5	19.1	303	16 Q7VE48	Q7VE48 prochloroco
26	291.5	19.0	302	16 Q7V3D3	Q7V3D3 prochloroco
27	291	19.0	602	3 Q9P854	Q9P854 phycomyces
28	291	19.0	602	3 Q9P877	Q9P877 phycomyces
29	287.5	18.7	287	16 Q7NHD5	Q7NHD5 bordetella
30	286.5	18.7	287	16 Q7W9C9	Q7W9C9 bordetella
31	286.5	18.7	287	16 Q7VYT8	Q7VYT8 bordetella
32	286	18.6	432	10 Q40166	Q40166 lycopersico
33	284.5	18.5	389	10 Q9AU04	Q9AU04 haematococc
34	279.5	18.2	228	10 Q9LLR8	Q9LLR8 tagetes ere
35	276.5	18.0	425	10 O04007	O04007 dunaliella
36	267.5	17.4	311	16 Q8DIM2	Q8DIM2 synechococc
37	266	17.3	310	16 Q8IVY5	Q8IVY5 anabaena sp
38	261.5	17.0	290	16 Q9JRU9	Q9JRU9 neisseria m
39	255.5	16.6	290	16 Q9JUF5	Q9JUF5 neisseria m
40	245	16.0	331	16 Q9RIX8	Q9RIX8 streptomyce
41	240	15.6	221	10 Q84N52	Q84N52 zea mays m
42	232	15.1	291	2 Q34288	Q34288 zymomonas m
43	232	15.1	342	16 Q93HP1	Q93HP1 streptomyce
44	231	15.0	153	10 Q84XR6	Q84XR6 citrus maxi
45	231	15.0	204	10 Q8L5H9	Q8L5H9 oryza sativ

## ALIGNMENTS

### RESULT 1

Q8GCS0	PRELIMINARY;	PRT;	296 AA.
ID	Q8GCS0		
AC	Q8GCS0;		
DT	01-MAR-2003 (TtEMBLrel. 23, Created)		
DT	01-MAR-2003 (TtEMBLrel. 23, Last sequence update)		
DT	01-OCT-2003 (TtEMBLrel. 25, Last annotation update)		
DB	Phytoene synthase.		
GN	CRFB.		
OS	Pantoea stewartii.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Pantoea.		
OX	NCBI_TaxID=56269;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 8200;		
RA	deSouza M.L., Kollmann S.R., Schroeder W.A.;		
RT	"Carotenoid Biosynthesis (WO 02/079395 A2).";		
RL	Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AV166713; AAN85600.1; -		
DR	GO; GO:0016740; F:transferase activity; IEA.		
DR	GO; GO:0009058; P:biosynthesis; IEA.		
DR	InterPro; IPR02060; Squ/phyt_synthase.		
DR	InterPro; IPR008949; Terpenoid_synthase.		
DR	Pham; PF00494; SQS_PSY; 1.		
DR	PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.		
DR	PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.		
SQ	SEQUENCE 296 AA; 33108 MW; B2294BI6D8513PC2 CRC64;		

Query Match 100.0%; Score 1535; DB 2; Length 296;

Best Local Similarity 100.0%; Pred. No. 4.2e-126;

Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAVGSKSPATASTLFPDKTRSVLMYAMCRCHDDVIDDQTLGFGHADQSSQMPERLQ 60

Db 1 MAVGSKSPATASTLFPDKTRSVLMYAMCRCHDDVIDDQTLGFGHADQSSQMPERLQ 60

Qy 61 LEWKTRQAVAGSOMHEPAFAFQEVAMAHDIAPAFHLEGFAMDVRETRYLTLDLTLR 120

61 LEMKTRQAVAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGFMVDRVRETRYLTLDLTLR 120  
121 YCHVAGVGLMMAQIMGVDRDNATLDRACDLGLAQLTNIARDIVDDAQVGCYLPESWL 180  
121 YCHVAGVGLMMAQIMGVDRDNATLDRACDLGLAQLTNIARDIVDDAQVGCYLPESWL 180  
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240  
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240  
241 KIGVKVEQAGKQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296  
241 KIGVKVEQAGKQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296

## RESULT 2

47846 PRELIMINARY; PRT; 296 AA.  
Q47846;  
01-NOV-1996 (TREMELrel. 01, Created)  
01-NOV-1996 (TREMELrel. 01, Last sequence update)  
01-OCT-2003 (TREMELrel. 25, Last annotation update)  
E Prephosphatase synthase.  
N CMTB.  
S Pantoea agglomerans.  
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
C Enterobacteriaceae; Pantoea.  
X NCBI\_TaxID=549;

Query Match 89.3%; Score 1371; DB 2; Length 296;  
Best Local Similarity 88.5%; Pred. No. 9.5e-112;  
Matches 262; Conservative 12; Mismatches 22; Indels 0; Gaps 0;  
1 MAVGSKSPATSTLFDKTRRSVLMYAWCRHCDVDVDDQTLGFGHADOPSSQMPERLQ 60  
1 MAVGSKSPATSKLFDKTRRSVLMYAWCRHCDVDVDDQTLGFGTQDPSLTPEQLMQ 60  
61 LEMKTRQAVAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGFMVDRVRETRYLTLDLTLR 120  
61 LEMKTRQAVAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGFMVDRVREASYIQLDNL 120  
121 YCHVAGVGLMMAQIMGVDRDNATLDRACDLGLAQLTNIARDIVDDAQVGCYLPESWL 180  
121 YCHVAGVGLMMAQIMGVDRDNATLDRACDLGLAQLTNIARDIVDDAQVGCYLPASWL 180  
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240  
181 ENEGLTKANYAAPENRQALSRIARLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240  
241 KIGVKVEQAGKQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296  
241 KIGVKVEQAGKQAWDHROSTSTPEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296

## RESULT 3

Q8VUJ4 PRELIMINARY; PRT; 296 AA.  
ID Q8VUJ4  
AC Q8VUJ4  
DT 01-MAR-2002 (TREMELrel. 20, Created)  
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE CrtB protein.  
GN CMTB.  
OS Pantoea agglomerans pv. milletiae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pantoea.  
OX NCBI\_TaxID=182454;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Kamiyama H., Hirata R.;  
RT "Isolation and characterization of carotenoid biosynthesis genes from  
RT Pantoea agglomerans pv. milletiae Wist 801.";  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB076662; BAB79604.1; -;  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0009058; P:biosynthesis; IEA.  
DR InterPro; IPR002060; Squ/phyt\_synthase.  
DR InterPro; IPR008949; Terpenoid\_synth.  
DR Pfam; PF00494; SQS\_PSY; 1.  
DR PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.  
DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.  
SQ SEQUENCE 296 AA; 32696 MW; C2BFC3E0C3C3CE56 CRC64;

Query Match 81.0%; Score 1244; DB 2; Length 296;  
Best Local Similarity 82.1%; Pred. No. 1.2e-100;  
Matches 243; Conservative 12; Mismatches 41; Indels 0; Gaps 0;  
1 MAVGSKSPATSTLFDKTRRSVLMYAWCRHCDVDVDDQTLGFGHADOPSSQMPERLQ 60  
1 MEVGSKSPATSKLFGKTRRSVLMYAWCRHCDVDVDDQTLGFGSNDTPSLQSAEQRLAQ 60  
61 LEMKTRQAVAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGFMVDRVRETRYLTLDLTLR 120  
61 LEMKTRQAVAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGFMVDRVRETRYLTLDLTLR 120  
121 YCHVAGVGLMMAQIMGVDRDNATLDRACDLGLAQLTNIARDIVDDAQVGCYLPESWL 180  
121 YCHVAGVGLMMAQIMGVDRDNATLDRACDLGLAQLTNIARDIVDDAQVGCYLPATWL 180  
181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240  
181 AEEGLTKANYAAPENRQALSRIARLVREAEPPYVSSMAGLAQLPLRSAMAIATAKQYVR 240  
241 KIGVKVEQAGKQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296  
241 KIGMKVQVQAASQAWDROSTSTPEKLTLLLTASGQAVTSRVARHAPRSADLWQRPV 296

## RESULT 4

Q93C16 PRELIMINARY; PRT; 311 AA.  
ID Q93C16  
AC Q93C16;  
DT 01-DEC-2001 (TREMELrel. 19, Created)  
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Phytoene synthase.  
GN CMTB.  
OS Xanthobacter sp. (strain Py2).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Hyphomicrobiaceae; Xanthobacter.  
OX NCBI\_TaxID=78245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Py2;  
RA Larsen R.A., Metcalf W.W.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF408848; AAL02001.1; -;  
DR GO; GO:0016740; F:transferase activity; IEA.

GO; GO:0009058; P:biosynthesis; IEA.  
InterPro; IPR002060; Squ/phyt synthase.  
InterPro; IPR008949; Terpenoid synthase.  
Pfam; PF00494; SOS\_PSY; 1.  
PROSITE; PS01044; SQUALEN\_PHYTOEN\_SYN\_1; 1.  
PROSITE; PS01045; SQUALEN\_PHYTOEN\_SYN\_2; 1.  
SEQUENCE 311 AA; 33279 MW; 7752769786099CAA CRC64;  
Query Match 52.4%; Score 804; DB 2; Length 311;  
Best Local Similarity 54.9%; Pred. No. 4.4e-62;  
Matches 161; Conservative 37; Mismatches 93; Indels 2; Gaps 2;  
4 GSKSPATASTLFDKATRSVLMYAWCRHCDVDVDTGLGFHADQPSQMPERLQOLEM 63  
19 GSKSPAAARLFGPRMEDAYMLYAWCRHCDVDVDTGLGFHADQPSQMPERLQOLEM 77  
64 KTRQYAGSQVHEPAPAFQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYLTLDITLYCY 123  
78 ETRNAYRGAPSAHPAPAFQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYLTLDITLYCY 137  
124 HVAGVGLMAYQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWLEEE 183  
138 HVAGVGVVGMVXALMGARDEVVLDASDLGLGFQLTNIARDIVDDAQVGRCYLPESWLEEE 197  
184 GLTKANYAAPENRQALSRAGRLVREAEPPYVSSMAGLAQLPLRSAMATATAKQVYRKIG 243  
198 GVPERAIAAPERRAAVAGVVARLLDVAEPPYDQALIGMAALPFAAAAVGAARGVYRAIG 257  
244 VKVQAGQKQWHDHSTQSTABKLLTLLTASQAVTSRMK-TYPERPAHLWQRP 295  
258 LEVRKRGKAWDTRVSTSTAQAGFLAGLAQALATFRGAPPPRPAHLWTRP 310  
SULT 5  
KIXI  
Q9KIX1 PRELIMINARY; PRT; 335 AA.  
01-OCT-2000 (TREMELrel. 15, Created)  
01-OCT-2000 (TREMELrel. 15, Last sequence update)  
01-OCT-2003 (TREMELrel. 25, Last annotation update)  
Phytoene synthase.  
CRTB.  
Bradyrhizobium sp. CRS278.  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Bradyrhizobiaceae; Bradyrhizobium.  
NCBI\_TaxID=114615;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=ORS278;  
MEDLINE=20309720; PubMed=10851005;  
Hannibal L., Lorquin J., Argles d'Ortoli N., Garcia N.,  
Chaintreuil C., Masson-Boivin C., Dreyfus B., Giraud E.;  
"Isolation and characterization of the canthaxanthin biosynthesis  
genes from the photosynthetic bacterium Bradyrhizobium sp. strain  
ORS278";  
J. Bacteriol. 182:3850-3853 (2000).  
EMBL; AF218415; AAF78202.1; -  
GO; GO:0016740; P:transferase activity; IEA.  
GO; GO:0009058; P:biosynthesis; IEA.  
InterPro; IPR002060; Squ/phyt synthase.  
InterPro; IPR008949; Terpenoid synthase.  
Pfam; PF00494; SOS\_PSY; 1.  
PROSITE; PS01044; SQUALEN\_PHYTOEN\_SYN\_1; 1.  
PROSITE; PS01045; SQUALEN\_PHYTOEN\_SYN\_2; 1.  
SEQUENCE 335 AA; 36869 MW; 6010A5B12014FDD2 CRC64;  
Query Match 48.4%; Score 743.5; DB 2; Length 335;  
Best Local Similarity 51.7%; Pred. No. 9.6e-57;  
Matches 156; Conservative 40; Mismatches 87; Indels 19; Gaps 4;  
4 GSKSPATASTLFDKATRSVLMYAWCRHCDVDVDTGLGFHADQPSQMPERLQOLEM 61  
20 GSKSPAAASKLFDSTRASVHLLYAWCRHCDVDVDTGLGFHADQPSQMPERLQOLEM 77

QY 62 EMKTRQYAGSQVHEPAPAFQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYLTLDITLY 121  
DB 78 RDQTAQALEGAPMEDPVGGLQORVVQEHAIIPHHEVVELLDGFAMDVGREYETLSETLDY 137  
QY 122 CYHVAGVGLMAYQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWLE 181  
DB 138 CYHVAGVGVVGMVXALMGAREEATLDRADLGLALQLTNIARDIVDDAQVGRCYLPESWLE 197  
QY 182 SEGTLKANYAAPENRQALSRAGRLVREAEPPYVSSMAGLAQLPLRSAMATATAKQVYRK 241  
DB 198 EAGVPAEAEVAPERRAAVAGVVARLLDVAEPPYDQALIGMAALPFAAAAVGAARGVYRA 257  
QY 242 IGKVBQAGQKQWHDHSTQSTABKLLTLLTASQAVTSRMKTYPERPAHLWQ 293  
DB 258 IGREVKRGKAWDTRVSTSTAQAGFLAGLAQALATFRGAPPPRPAHLWTRP 310  
QY 294 RP 295  
DB 311 RP 312  
RESULT 6  
P94789 PRELIMINARY; PRT; 303 AA.  
AC P94789  
DT 01-MAY-1997 (TREMELrel. 03, Created)  
DT 01-MAY-1997 (TREMELrel. 03, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Phytoene synthase.  
GN CRTB.  
OS Flavobacterium sp. ATCC 21588.  
OC Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales;  
OC Flavobacteriaceae; Flavobacterium.  
OX NCBI\_TaxID=50286;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP STRAIN=R1534;  
RX MEDLINE=97186694; PubMed=9034310;  
RA Pasamontes L., Hug D., Tessier M., Hohmann H.P., Schierle J.,  
van Loon A.P.;  
"Isolation and characterization of the carotenoid biosynthesis genes  
of Flavobacterium sp. strain R1534";  
RL Gene 185:35-41 (1997).  
DR EMBL; U62808; AAC44849.1; -  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0009058; P:biosynthesis; IEA.  
DR InterPro; IPR002060; Squ/phyt synthase.  
DR InterPro; IPR008949; Terpenoid synthase.  
DR Pfam; PF00494; SOS\_PSY; 1.  
DR PROSITE; PS01044; SQUALEN\_PHYTOEN\_SYN\_1; 1.  
DR PROSITE; PS01045; SQUALEN\_PHYTOEN\_SYN\_2; 1.  
SQ SEQUENCE 303 AA; 32616 MW; 535CC680B4D9D0 CRC64;  
Query Match 44.7%; Score 686.5; DB 2; Length 303;  
Best Local Similarity 48.7%; Pred. No. 8.2e-52;  
Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;  
QY 1 MAVSKSPATASTLFDKATRSVLMYAWCRHCDVDVDTGLGFHADQPSQMPERLQOLEM 60  
DB 12 IAQGSQSPAAAKLMPFGIREDTVMYAWCRHCDVDVDTGLGFHADQPSQMPERLQOLEM 69  
QY 61 LEMKTRQYAGSQVHEPAPAFQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYLT 115  
DB 70 LEADTLAA-----LHEDGPMSPPPAALRQVARRHDFPDLMPMDLIEGFAMDVARRRSL 124  
QY 116 DDTLYCYHVAGVGLMAYQIMGVNRNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYL 175  
DB 125 DDVLEYSHVAGVGVVGMVXALMGAREEATLDRADLGLALQLTNIARDIVDDAQVGRCYL 184  
QY 176 PESWLEEGTLKANYAAPENRQALSRAGRLVREAEPPYVSSMAGLAQLPLRSAMATATA 235  
DB 185 PADWLAEG---ATVEGVPFSDALYSVIRLLDAEPPYASARQGLPHLPFCAMWSTAAA 241



236 KQYKICGVQVQAGQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 295  
 242 LRIYALGTIRQGGPEAYRQISTSKAKIGLLARGGLDAAASRLRGHISRDGLWTRP 301

## RESULT 7

9RLH3 PRELIMINARY; PRT; 304 AA.  
 C Q9RLH3; 01-MAY-2000 (TrEMBLrel. 13, Created)  
 T 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 I 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 E Phytoene synthase.  
 N CRFB.  
 S Paracoccus marcusii.  
 C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
 C Rhodobacteraceae; Paracoccus.  
 X NCBI\_TaxID=59779;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C STRAIN=MHI;  
 A "Carotenoid biosynthesis genes in the bacterium Paracoccus marcusii  
 T MHI";  
 L Submitted (SEP-1997) to the EMBL/GenBank/DBSJ databases.  
 R ENBL; Y15112; CAB56063.1; --  
 R GO; GO:0016740; P:transferase activity; IEA.  
 R GO; GO:0009058; P:biosynthesis; IEA.  
 R InterPro; IPR002060; Squ/phyt synthase.  
 R InterPro; IPR008949; Terpenoid\_synth.  
 R Pfam; PF00494; SOS\_PSY; 1.  
 Q SEQUENCE 304 AA; 33185 MW; 0C0BA7CEDC30828A CRC64;

Query Match 42.1%; Score 645.5; DB 2; Length 304;  
 Best Local Similarity 46.9%; Pred. No. 3.2e-48;  
 Matches 137; Conservative 35; Mismatches 115; Indels 5; Gaps 2;

Y 4 GSKSPATATLFDKTRRSVLMYAWCRHCDVIDDQTLGFHADQFSSQMPERLQOLEM 63  
 b 15 GSQSPATAXLPPGIRDTVWLYAWCRHADDVIDQGLGRPEAVND--PQARLDGLRA 72  
 Y 64 KTRQAYAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGFPAMDVRETRVLTDDTLRYCY 123  
 b 73 DTLAALQGGPVPPTPPFAALRAVARRHDPQAWPMDLIEGFAIDVEARDYRTLDDVLEYSY 132  
 Y 124 HVAGVGLMMAQIMGVDMATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWLEEE 183  
 b 133 HVAGILGVMMARVGVDRDHPVLDRACDLGLAFQLTNIARDIVDIYARIGRCYLPGLDLOHA 192  
 Y 184 GLTKANYAAPENRQALSRIAGRLVREAPFYVSSMAGLAQLPLRSAMAIATAKOVYRKIG 243  
 b 193 G--ARVDGVPSPSLYTVILRLDAAEPYPSARVGLADUPPCAWSIPAAFRYIYRAIG 249  
 Y 244 VKVQAGQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 295  
 b 250 LPIRKGPEAYRQISTSKAKIGLLGIGGWMDVADHACRGSGVSRQDLWTRP 301

## RESULT 8

3KXZ7 PRELIMINARY; PRT; 346 AA.  
 C Q8KZ27; 01-OCT-2002 (TrEMBLrel. 22, Created)  
 T 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 I 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 E Phytoene synthase.  
 N CRFB.  
 S uncultured proteobacterium.  
 C Bacteria; Proteobacteria; environmental samples.  
 X NCBI\_TaxID=153809;  
 N [1]  
 P SEQUENCE FROM N.A.

MEDLINE=21822632; PubMed=11832943;  
 RA Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,  
 RA Hamada T., Eisen J.A., Fraser C.M., DeLong E.F.;  
 RT "Unsuspected diversity among marine aerobic anoxygenic phototrophs.";  
 RL Nature 415:630-633(2002).

DR EXBL; AE008921; AAM48647.1; --  
 DR GO; GO:0016740; P:transferase activity; IEA.  
 DR GO; GO:0009058; P:biosynthesis; IEA.  
 DR InterPro; IPR002060; Squ/phyt synthase.  
 DR InterPro; IPR008949; Terpenoid\_synth.  
 DR Pfam; PF00494; SOS\_PSY; 1.  
 DR PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.  
 DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.  
 SQ SEQUENCE 346 AA; 37857 MW; 13A5C6B82FAC1237 CRC64;

Query Match 29.1%; Score 446.5; DB 2; Length 346;  
 Best Local Similarity 38.0%; Pred. No. 1e-30;  
 Matches 115; Conservative 35; Mismatches 120; Indels 33; Gaps 5;

QY 4 GSKSPATATLFDKTRRSVLMYAWCRHCDVIDDQTLGFHADQFSSQMPERLQOLEM 63  
 Db 17 GSLSFHAASKLLPASVRDPALALYAFCLADDEVDE-----GQNKTRAVIELQE 65  
 QY 64 KTRQAYAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGFPAMDVRETRVLTDDTLRYCY 123  
 Db 66 RLALVYAGPRNAPEDRAPASVVEDEFEMPAALPEALLEGLANDEHRYSSLSDLRGYCA 125  
 QY 124 HVAGVGLMMAQIMGVDMATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWLEEE 183  
 Db 126 RVASAVGAMCVLMRVDRDADALARACDLGVAMQLTNIARDVGEDARAGRIVLEWIDAE 185  
 QY 184 GL-----TKANYAAPENRQALSRIAGRLVREAPFYVSSMAGLAQLPLRSAMAIATAKOV 238  
 Db 186 GLDPQVLSVTATPE-----LRMWKLLSEHAALYVRSEAGVAALPLNARTGIYAAYI 241  
 QY 239 YRKIGVYKQAGQAWDHROSTSTAEKLTLL-----TASGQAVTSRMKTYPPRPAHLWQ 293  
 Db 242 YDAIGQAVARNHYDSITHRGRTTKAQKALLAKSLRTAAGLVW-----PRSPVLVA 293  
 QY 294 RPI 296  
 Db 294 RPL 296

## RESULT 9

Q8RTY0 PRELIMINARY; PRT; 379 AA.  
 ID Q8RTY0  
 AC Q8RTY0; 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Phytoene synthase.  
 GN CRFB.  
 OS uncultured proteobacterium.  
 OC Bacteria; Proteobacteria; environmental samples.  
 OX NCBI\_TaxID=153809;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA MEDLINE=21822632; PubMed=11832943;  
 RA Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,  
 RA Hamada T., Eisen J.A., Fraser C.M., DeLong E.F.;  
 RT "Unsuspected diversity among marine aerobic anoxygenic phototrophs.";  
 RL Nature 415:630-633(2002).

DR EXBL; AE008919; AAL76346.1; --  
 DR GO; GO:0016740; P:transferase activity; IEA.  
 DR GO; GO:0009058; P:biosynthesis; IEA.  
 DR InterPro; IPR002060; Squ/phyt synthase.  
 DR InterPro; IPR008949; Terpenoid\_synth.  
 DR Pfam; PF00494; SOS\_PSY; 1.  
 DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.  
 SQ SEQUENCE 379 AA; 41383 MW; 3AA1605561C3CC6D CRC64;

Query Match 26.3%; Score 403; DB 2; Length 379;

```

RT Rubrivivax gelatinosus.";
RL Biochemistry 38:15238-15244 (1999).
EMBL AB034704; BAA94048.1; -.
PTR; T50895; T50895.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009058; P:biogenesis; IEA.
DR InterPro; IPR002060; Squ/phyt_synthese.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00494; SOS_PSY; 1.
DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
SQ SEQUENCE 344 AA; 37551 MW; CCC7F8AD52AA570 CRC64;

Query Watch 25.8%; Score 395.5; DB 2; Length 344;
Best Local Similarity 36.2%; Pred.No.2.9e-26;
Matches 106; Conservative 32; Mismatches 140; Indels 15; Gaps 3;

QY 1 MAVSKSFATASTIFDAKTRRSVLMYAVMCRHCDVDIDDTLGFHADQPSOMPQRLEQQ 60
DB 18 MEGGSKFFAASLLLPQVRTPATALVACRVADDAVD-----LSGDPHAAMDE 66

QY 61 LEMKTRQYAGSQMHSEPAFAAFOEVAMAHDIAPAYAFHLEGFAMDVRETRYLYLDDTLR 120
DB 67 LTRTRIDAVYAGTPAPIAADRALASTVHRYGVPVRVLLDALLEGFLWDADGRYDTIADVEA 126

QY 121 CYCHVAGVGLMMAQINGVDNATLDRACDLGLAFOLNTIARDIVDDAQGRCYLPESWL 180
DB 127 YCARVAGTVGAMALINGVASPQALARAELGVAMQFTNIARDVGEDARNGELYPREWL 186

QY 181 EEEGL-TKANYAAPENRQALSRAGRLVREASEPYVSSWAGLAQLPLRSAMAIATAKOVY 239
DB 187 VEAGLDVDAMLQNVHCPEVQAVTRLLRAADBELYERSEHGIAALPRCPRPAIRAARLVY 246

QY 240 RKGIVKVEQAGKQAWDHROSTSTAETKLITLLTASGOAVTSRMKTY---PRRPA 289
DB 247 AEIGMLERGLDSVNRHVVPARRKAALMARASAAFNTPGRAYISMPPLPA 299

RESULT 11
QSRW07
ID QSRW07 PRELIMINARY; PRT; 325 AA.
AC QSRW07.
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DE Phytoene synthase.
GN DR0862.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
(1)
RN RN
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.P., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Hatt D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Yamathavan J.S., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.J., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RA "Genome sequence of the radioresistant bacterium Deinococcus
RA radiodurans R1.";
RL Science 286:1571-1577(1999).
EMBL AB001940; AAF10440.1; -.
DR PIR; D75466; D75466.
DR TIGR; DR0862; -.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009058; P:biogenesis; IEA.
DR InterPro; IPR002060; Squ/phyt_synthese.
DR InterPro; IPR008949; Terpenoid_synth.
DR Pfam; PF00494; SOS_PSY; 1.
DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.

```

```
QW Complete proteome.
SQ SEQUENCE 325 AA; 35972 MW; 52BJADB66C65853F CRC64;

Query Match
Best Local Similarity 23.6%; Score 362; DB 16; Length 325;
Matches 103; Conservative 37; Mismatches 111; Indels 24; Gaps 8;

Y 5 SKSPATASTLFDKATRSVLMLVANCRCDDVDVDTGLFHADQPSQMPQRLQLEMK 64
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 50 SKTFYGLSQFSPPEAAVAVVAVVAAACRAGDDVDEAGNG---DR-----ERELREWSR 100
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 65 TRQAYAGSQMHGEPAPAFQEVANAHDIAPAYAFDHL-EGFAMDVRETRYLTLDITLVCY 123
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 101 IDAFAGQADDPDLSTALAWAGRYAI-PSHSAFELHGLNMDLRGHEYRDMDDLLLYCR 159
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 124 HVAGVGLMMAQIMGVADN-ATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWLEE 182
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 160 RVAGVGFVAPISGYGGAATLNDALQLQAMQLTNILRDVGEDLTRGVYLPQSLLDE 219
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 183 EGLTKA-----NYAAPENRQALSRAGRLVREAPYVYSSMAGLAQLPLRSWAIAT 234
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 220 YGSRAALERWGGQEPUSPAYRALMTHLGG-LAR3---WYAAGRAGIPQLDGRGLAVLT 275
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 235 AKQYRKIGVKVQAGKQAWDHRQSTSTAEKLTLL 269
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 276 AARAYEGILDLERAGVDNFGRAYVSGREKULML 310
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
D OS0567 PRELIMINARY; PRT; 343 AA.
C OS0567;
T 01-JUN-1998 (T-EMBLrel. 06, Created)
T 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
E Phytoene synthase.
N CRTB.
S Rhodocyclus gelatinosus (Rhodospseudomonas gelatinosa).
X Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
X Comamonadaceae; Rubrivivax.
X NCB1_TaxID=28068;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=S1;
X MEDLINE=974747582; PubMed=9303322;
A Ouchane S., Picaud M., Vernotte C., Astier C.;
T "Photooxidative stress stimulates illegitimate recombination and
T mutability in carotenoid-less mutants of Rubrivivax gelatinosus.";
L EMBO J. 16:4777-4787(1997).
R EMBL; AY234384; AAB87738.1; -.
R GO; GO:0016740; F:transferase activity; IEA.
R GO; GO:0009058; P:biosynthesis; IEA.
R InterPro; IPR002060; Squ/phyt synthase.
R InterPro; IPR008949; Terpenoid_synth.
R Pfam; PF00434; SQS_PSY; 1.
R PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
Q SEQUENCE 343 AA; 37509 MW; C7CA4F5229B9CBFAF CRC64;

Query Match
Best Local Similarity 22.9%; Score 351; DB 2; Length 343;
Matches 102; Conservative 31; Mismatches 123; Indels 22; Gaps 6;

Y 1 MAVGSKSFATASTLFDKATRSVLMLVANCRCDDVDVDTGLFHADQPSQMPQRLQ 60
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 18 MRGSGSKSFFAASLLLPQVRAPATALYAFCRVADDAVD-----LSDGPEANMAE 66
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 61 LEMKTRQAYAGSQMHGEPAPAFQEVANAHDI-----APAYAFDH-LEGFAMDVRETRYLT 115
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 67 LRTFELDGVYAGT-----PAPIA-ADRALACTVHYGVPRVILLALLEGFLWDADGRYDTI 121
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Y 116 DTLRYCYHVAGVGLMMAQIMGVADNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYL 175
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 122 ADVEAYGARVAGTVGAMALIMGVRSFQALARACELGVAMQFTNI-ARDVGEDARNGRLYL 181
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Query Match
Best Local Similarity 20.4%; Score 313.5; DB 10; Length 422;
Matches 80; Conservative 59; Mismatches 125; Indels 19; Gaps 6;

QY 5 SKSPATASTLFDKATRSVLMLVANCRCDDVDVDTGLFHADQPSQMPQRLQLEMK 64
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 138 AKTFYGLTMTPTERQKAIWAIYVWCRRTDELVDGPN-----ASHITPKALDRWESR 189
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 65 TRQAYAGSQMHGEPAPAFQEVANAH--DIAPAYAFDHLEGFAMDVRETRYLTLDITLYC 122
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 190 LEDLFNG-RPFDMLDAALSDTVSKFPVDIQPFK--DMIEGMNDLRKSKYQNFDELYLC 246
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 123 YHVAGVGLMMAQIMGV--RDNATLD---RACDLGLAFQLTNIARDIVDDAQVGRCYLP 176
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 247 YYVAGTVGLMSPVIMGIDPESQATTVESVYNALALGIANQLNILLRDVGEDARRGVYLP 306
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 177 ESWLEBEGLTKANYAAPENRQALSRAGRLVREAPYVYSSMAGLAQLPLRSWAIATK 236
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 307 QDELAQAGLSDEDFSEKVTDKWRYPMKQIKRARSFFDEAQQGVTLSSASGRWPVSASL 366
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 237 QVYRKIGVKVQAGKQAWDHRQSTSTAEKLTLLLTASGOAVTS 279
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
b 367 FLTRQILDEIANDYNNFTKRYAVSKSKLVSPLVAYAKSLVS 409
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
QYXGW7
ID QXKGW7 PRELIMINARY; PRT; 436 AA.
AC QXKGW7;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Phytoene synthase.
```

Citrus paradisi (Grapefruit).  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
eurosids II; Sapindales; Rutaceae; Citrus.  
NCBI\_TaxID=37656;  
[1]  
SEQUENCE FROM N.A.  
Costa M.C., Moreira C.D., Melton J.R., Otoni W.C., Moore G.A.;  
"Developmental expression of carotenoid genes in Citrus."  
Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
EMBL; AF152892; AAD38051.2;  
GO: GO:0016740; P:transferase activity; IEA.  
GO: GO:0009058; P:biosynthesis; IEA.  
InterPro; IPR002060; Squ/pyr synthase.  
Pfam; PF00494; SOS\_PSY; 1.  
PROSITE; PS01044; SQUALEN\_PHYTOEN SYN 1; 1.  
PROSITE; PS01045; SQUALEN\_PHYTOEN SYN 2; 1.  
SEQUENCE 436 AA; 49360 MW; 87F7C8D798FA88B1 CRC64;  
Query Match 20.4%; Score 313.5; DB 10; Length 436;  
Best Local Similarity 28.1%; Pred. No. 6.1e-19;  
Matches 81; Conservative 56; Mismatches 132; Indels 19; Gaps 5;  
5 SKSFATASTLPDAKTRRSVLMYAKCRHCDVIDDQTLGFHADQPSSQMPBQRLQQLQLEMK 64  
156 AKTFYLGTLMTSERRALWAIYVWCRFTDELVDGPN-----ASHITPDLRWESR 207  
65 TRQAVAGSQMHPEAFAPQEVANAH--DIAPAYAFDHLGFMADVRETRYTLDDTLRYC 122  
208 LEDLFRG-QPFDMLDAALSDTVTKFPVDIQPFR--DMIEGVRMDLRKSRVKNFDELYLC 264  
123 YHVAGVGLMAQIMGVDRN-----ATLDRACDLGLAFQLTNIARDIVDDAQVGRCLYP 176  
265 YVAGTVGLMSVPVNGIAPDSQATTESVYNHALGIANQLTNILRDVGDGRPRGYLP 324  
177 ESWLEEEGLTKANYAAPENRQALSRIAGRLVREAPYVYSSMAGLAQLPLSAWAIATAK 236  
325 QDLAQAGLSDDIDPAGEVTIKWRNFKNQIKRSMFFDMAENGVTLSSEASRWFVWASL 384  
237 QYRKIGVYKVGQAGQKQMDHRQSTSTAEKLTLLLTASQAVTSRMKTY 284  
385 LLYRQILDEIANDYNNFTKRCACVSKAKXKXIALPIAYAKSLLRPSRIY 432  
RESULT 15  
JU419 PRELIMINARY; PRF; 303 AA.  
Q7U419;  
01-OCT-2003 (TrEMBLrel. 25, Created)  
01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
Phytoene synthases (EC 2.5.1.-).  
CRTL, PYS OR SYNW2256.  
Synecococcus sp. (strain WH9102).  
C Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
NCBI\_TaxID=84588;  
[1]  
SEQUENCE FROM N.A.  
MEDLINE=2825697; PubMed=12917641;  
A Palenik B., Brahmasha B., Larimer F.W., Land M., Hauser L., Chain P.,  
Lamerdin J., Regala W., Allen E.E., McCarren J., Paulsen I.,  
Buiresne A., Partensky F., Webb E.A., Waterbury J.;  
I "The genome of a motile marine Synecococcus."  
L Nature 424:1037-1042(2003).  
R EMBL; BX569695; CAE08771.1; --  
W Transferase; Complete proteome.  
Q SEQUENCE 303 AA; 34984 MW; 05527FA419F80005 CRC64;  
Query Match 20.2%; Score 310.5; DB 16; Length 303;  
Best Local Similarity 31.8%; Pred. No. 6.8e-19;  
Matches 87; Conservative 43; Mismatches 125; Indels 19; Gaps 6;

QY 5 SKSFATASTLPDAKTRRSVLMYAKCRHCDVIDDQTLGFHADQPSSQMPBQRLQQLQLEMK 64  
DB 24 AKTFYLGTLMTSERRALWAIYVWCRFTDELVDGPN-----ASHITPDLRWESR 207  
QY 65 TRQAVAGSQMHPEAFAPQEVANAH--DIAPAYAFDHLGFMADVRETRYTLDDTLRYC 123  
DB 80 LEDLFRG-QPFDMLDAALSDTVTKFPVDIQPFR--DMIEGVRMDLRKSRVKNFDELYLC 264  
QY 124 YHVAGVGLMAQIMGVDRN-----ATLDRACDLGLAFQLTNIARDIVDDAQVGR 172  
DB 138 RVAGTVGLMSVPVNGIAPDSQATTESVYNHALGIANQLTNILRDVGDGRGR 196  
QY 173 ESWLEEEGLTKANYAAPENRQALSRIAGRLVREAPYVYSSMAGLAQLPLSAWAI 232  
DB 197 QYRKIGVYKVGQAGQKQMDHRQSTSTAEKLTLLLTASQAVTSRMKTY 284  
QY 233 QDLAQAGLSDDIDPAGEVTIKWRNFKNQIKRSMFFDMAENGVTLSSEASRWFVWASL 384  
DB 257 LLYRQILDEIANDYNNFTKRCACVSKAKXKXIALPIAYAKSLLRPSRIY 432  
Search completed: February 29, 2004, 14:51:04  
Job time : 27.3391 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

on on: February 29, 2004, 14:27:18 ; Search time 4.98876 Seconds  
(without alignments)  
3089.496 Million cell updates/sec

title: US-09-941-947A-34

effect score: 1535  
sequence: 1 NAVGSKSPATSTLFDKTR.....VTSRMKTYPPRPAHLWQRPPI 296

scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1368	89.1	309	1 CRTB_PANAN	P21683 pantoeca ana
2	1028	67.0	309	1 CRTB_BRWHE	P22872 erwinia her
3	648	42.2	301	1 CRTB_AGRAU	P54975 agrobacteri
4	394	25.7	355	1 CRTB_RHOSH	P54905 rhodobacter
5	393.5	25.6	339	1 CRTB_RHOCA	P17056 rhodobacter
6	328.5	21.4	423	1 PSY_NARPS	P53797 nartissus p
7	309.5	20.2	419	1 PSY_CAPAN	P37272 capsicum an
8	307.5	20.0	310	1 PSY2_LYCES	P37273 lycopersico
9	307.5	20.0	398	1 PSY_DAUCA	Q98808 daucus caro
10	307	20.0	422	1 PSY_ARATH	P37271 arabidopsis
11	302.5	19.7	410	1 PSY_MAIZE	P49085 zea mays [m
12	300.5	19.6	412	1 PSY1_LYCES	P08196 lycopersico
13	297	19.3	602	1 PSY_NEUCR	P37295 neurospora
14	286	18.6	422	1 PSY_CUCME	P49293 cucumis mel
15	280	18.2	309	1 CRTY_SPIPL	O07333 spirulina p
16	273.5	17.8	614	1 LCPS_RHIRA	Q94066 rhizomucor
17	273	17.8	289	1 CRTB_THERH	P37270 thermus the
18	265.5	17.3	337	1 CRTB_SVNY3	P37294 synechocyst
19	258	16.8	307	1 CRTB_SVNP7	P37269 synechocyst
20	233.5	15.2	342	1 CRTB_STRGR	P54977 streptomyce
21	217.5	14.2	302	1 CRTB_WYCTU	O50728 mycobacteri
22	207	13.5	279	1 Y4AC_RHISN	P55350 rhizobium s
23	171.5	11.2	417	1 FDFT_HUMAN	P37268 homo sapien
24	163.5	10.7	416	1 FDFT_MOUSE	P37398 mus musculu
25	159.5	10.4	416	1 FDFT_RAT	Q02769 rattus norv
26	159	10.4	292	1 Y4AD_RHISN	P55351 rhizobium s
27	149	9.7	443	1 FDFT_CANGA	Q98926 candida gla
28	145.5	9.5	448	1 FDFT_CANAL	P37889 candida alb
29	130.5	8.5	460	1 FDFT_SCHPO	P36596 schizosach
30	124	8.1	444	1 FDFT_YEAST	P29704 nicotiana b
31	116.5	7.6	411	1 FDFT_NICEE	P53800 nicotiana b
32	116	7.6	410	1 FDFT_ARATH	P53799 arabidopsis
33	109.5	7.1	572	1 FDFT_USTMA	Q92459 ustilago ma

## ALIGNMENTS

### RESULT 1

ID	CRTB_PANAN	STANDARD;	PRT;	309 AA.
AC	P21683;	1991 (Rel. 18, Created)		
DT	01-MAY-1991	(Rel. 42, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Phytoene synthase (EC 2.5.1.-)			
GN	CRTB.			
OS	Pantoeca ananas (Erwinia uredovora).			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Pantoeca.			
OX	NCBI_TaxID=553;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=20D3;			
EX	MEDLINE=91072214; PubMed=2254247;			
RA	Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,			
RA	Nakamura K., Harashima K.;			
RT	"Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway			
RT	by functional analysis of gene products expressed in Escherichia			
RT	coli.";			
RL	J. Bacteriol. 172:6704-6712(1990).			
RN	[2]			
RP	REVISIONS TO N-TERMINUS.			
RA	Misawa N., Nakagawa M., Kobayashi K., Yamano S., Izawa Y.,			
RA	Nakamura K., Harashima K.;			
RL	Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.			
CC	-I- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to			
CC	phytoene.			
CC	-I- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +			
CC	prephytoene diphosphate.			
CC	-I- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +			
CC	phytoene.			
CC	-I- PATHWAY: Carotenoid biosynthesis.			
CC	-I- SIMILARITY: Belongs to the phytoene/squalene synthetase family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; D90087; BAA14128.2; --			
DR	PIR; E37802; E37802.			
DR	InterPro; IPR002060; Squ/phyt_synth.			
DR	InterPro; IPR008949; Terpenoid_synth.			
DR	Pfam; PF00494; SQS_PSV; 1.			
DR	PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.			
DR	PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.			
KW	Multifunctional enzyme; Carotenoid biosynthesis; Transferase.			
SQ	SEQUENCE 309 AA; 34472 MW; 9AA381A7376BBFC9 CRC64;			
Query Match	89.1%;	Score 1368;	DB 1;	Length 309;

Best Local Similarity 88.9%; Pred. No. 5.6e-115;  
Matches 263; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

1 MAVGSKSFATASTLFAKTRRSVLMLYAWCRHCDVDDVDDQTLGFFADQPSQWPQRLLQ 60  
14 MAVGSKSFATSKLFAKTRRSVLMLYAWCRHCDVDDVDDQTLGFFADQPSQWPQRLLQ 73  
61 LEMKTRQAYAGSQHMEPAPAFQEVAMAHDIAPAYAFDHLGFMNDVRETRVLTDDTLR 120  
74 LEMKTRQAYAGSQHMEPAPAFQEVAMAHDIAPAYAFDHLGFMNDVRETRVLTDDTLR 133  
121 YCVHAGVGLMAQINGVNDNATLDRACDLGLAFQLTWIAARDIVDDAQVGRCYLPESWL 180  
134 YCVHAGVGLMAQINGVNDNATLDRACDLGLAFQLTWIAARDIVDDAQVGRCYLPASWL 193  
181 EESGLTKANYAAPENQALSRIRAGLVRRAEPYVSSMAGLAQLPLRSAWATAKQVYR 240  
194 EHEGLNKENTAAPENQALSRIRAGLVRRAEPYVSSMAGLAQLPLRSAWATAKQVYR 253  
241 KIGVKVEQAGQAWDRQSTSTAEKLTLLLTASQAVTGRMKTYPPRPAHLWOPRI 296  
254 KIGVKVEQAGQAWDRQSTSTAEKLTLLLTASQAVTGRMKTYPPRPAHLWOPRI 309

RESULT 2  
RTB\_ERWHE STANDARD; PRT; 309 AA.  
C P22872, 1991 (Rel. 19, Created)  
01-AUG-1991 (Rel. 19, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
Phytoene synthase (EC 2.5.1.-).  
Erwinia herbicola.  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Pantoea.  
NCBI\_TaxID=549;  
SEQUENCE FROM N.A.  
STRAIN=EH010;  
MEDLINE=91088634; PubMed=2263648;  
Armstrong G.A., Alberti M., Hearst J.E.;  
"Conserved enzymes mediate the early reactions of carotenoid  
biosynthesis in nonphotosynthetic and photosynthetic prokaryotes."  
Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979 (1990).  
[2]  
FUNCTION.  
RP MEDLINE=92357711; PubMed=1495965;  
Math S.K., Hearst J.E., Poulter C.D.;  
"The crtB gene in Erwinia herbicola encodes geranylgeranyl  
diphosphate synthase."  
Proc. Natl. Acad. Sci. U.S.A. 89:6761-6764 (1992).  
RL -!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to  
phytoene.  
CC -!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +  
prephytoene diphosphate.  
CC -!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +  
phytoene.  
CC -!- PATHWAY: Carotenoid biosynthesis.  
CC -!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; M38423; AAA24821.1; -  
EMBL; M87280; AAA64582.1; -  
PIR; E39273; E39273.  
InterPro; IPR002060; Squ/phyt\_synthes.

InterPro; IPR008949; Terpenoid\_synth.  
DR Pfam; PF00444; SOS\_PSY; 1.  
DR PROSITE; PS01044; SQUALEN PHYTOEN SYN\_1; 1.  
DR PROSITE; PS01045; SQUALEN PHYTOEN SYN\_2; 1.  
KW Multifunctional enzyme; Carotenoid biosynthesis; Transferase.  
SQ SEQUENCE 309 AA; 34123 MW; B64235C9E0C06F38 CRC64;

Query Match 67.0%; Score 1028; DB 1; Length 309;  
Best Local Similarity 64.9%; Pred. No. 1.4e-64;  
Matches 192; Conservative 36; Mismatches 68; Indels 0; Gaps 0;

1 MAVGSKSFATASTLFAKTRRSVLMLYAWCRHCDVDDVDDQTLGFFADQPSQWPQRLLQ 60  
14 MAVGSKSFATSKLFAKTRRSVLMLYAWCRHCDVDDVDDQTLGFFADQPSQWPQRLLQ 73  
61 LEMKTRQAYAGSQHMEPAPAFQEVAMAHDIAPAYAFDHLGFMNDVRETRVLTDDTLR 120  
74 LEMKTRQAYAGSQHMEPAPAFQEVAMAHDIAPAYAFDHLGFMNDVRETRVLTDDTLR 133  
121 YCVHAGVGLMAQINGVNDNATLDRACDLGLAFQLTWIAARDIVDDAQVGRCYLPESWL 180  
134 YCVHAGVGLMAQINGVNDNATLDRACDLGLAFQLTWIAARDIVDDAQVGRCYLPASWL 193  
181 EESGLTKANYAAPENQALSRIRAGLVRRAEPYVSSMAGLAQLPLRSAWATAKQVYR 240  
194 ODAGLTPTENYAAPENQALSRIRAGLVRRAEPYVSSMAGLAQLPLRSAWATAKQVYR 253  
241 KIGVKVEQAGQAWDRQSTSTAEKLTLLLTASQAVTGRMKTYPPRPAHLWOPRI 296  
254 EIGIKVKAAGSADWRQSTSTAEKLTLLLTASQAVTGRMKTYPPRPAHLWOPRI 309

RESULT 3  
CRTB\_AGRAU STANDARD; PRT; 301 AA.  
ID CRTB\_AGRAU  
AC P54975;  
DT 01-OCT-1996 (Rel. 34, Created)  
01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Phytoene synthase (EC 2.5.1.-).  
GN CRTB.  
OS Agrobacterium aurantiacum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
OX NCBI\_TaxID=44155;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96062243; PubMed=7592436;  
RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiura S., Saito T.,  
Ohtani T., Miki W.;  
"Structure and functional analysis of a marine bacterial carotenoid  
biosynthesis gene cluster and astaxanthin biosynthetic pathway  
proposed at the gene level."  
J. Bacteriol. 177:6575-6584 (1995).  
RL -!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate  
to phytoene.  
CC -!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +  
prephytoene diphosphate.  
CC -!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +  
phytoene.  
CC -!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin  
biosynthetic pathway.  
CC -!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; D58420; BAA09595.1; -  
InterPro; IPR002060; Squ/phyt\_synthes.



```

C Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
C Rhodobacteraceae; Rhodobacter.
X NCBI_TaxID=1061;
N SEQUENCE FROM N.A.
X STRAIN=SB11003 / St. Louis, and BEC404;
MEDLINE=89313663; PubMed=2747617;
A Armstrong G.A., Alberti M., Leach P., Hearst J.E.;
I "Nucleotide sequence, organization, and nature of the protein
I products of the carotenoid biosynthesis gene cluster of Rhodobacter
I capsulatus.";
L Mol. Gen. Genet. 216:254-268 (1989).
C -1- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
C phytoene.
C -1- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
C prephytoene diphosphate.
C -1- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
C phytoene.
C -1- PATHWAY: Carotenoid and chlorophyll biosynthesis.
C -1- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
C EMBL; X52291; CAA36534.1; -.
C PIR; S04403; S04403.
C InterPro; IPR002060; Sm/phyt_synthse.
C InterPro; IPR008949; Terpenoid_synth.
C Pfam; PF00494; SQS_PSI; 1.
C PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
C PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
C Multifunctional enzyme; Carotenoid biosynthesis; Transferase;
C Chlorophyll biosynthesis.
C SEQUENCE 339 AA; 37301 MW; 6D7052994934BF4F CRC64;

Query Match 25.6%; Score 393.5; DB 1; Length 339;
Best Local Similarity 35.6%; Pred. No. 9.1e-28;
Matches 103; Conservative 44; Mismatches 125; Indels 17; Gaps 6;

Y 4 GSKSFATATLFDKTRRSVLMVYAWCRHCDVDVDDQTLGPHADQPSQMPQRLQOLEM 63
Y 17 GSYSFHAASRVLPARVEDPALALYAFCRVADDEVE--VGAPDKAAAVL-----KLGD 68
Y 64 KTRQAYAGSQMHPEP---AFAA-FOEVAMAHDIAPAYAFPHLGFAMDVRETRYLTLDL 119
Y 69 RLEDIYAGRPRNAPSDFRAFAVVEEFEMPRLPEAL----LEGFADWABGRWTHLSVQ 124
Y 120 RYCVHGVGVGLMMAQIMGVYRDNATLDRCADGLAQTLNIAARDIVDDAQVGRCYLP 179
Y 125 AYSARVAAAAGMVCMLRVNPNALARAADGLAQMSNIARDVGEDARGLRFLPTDM 184
Y 180 LEEBGL-TKANYAAPENRQALSRIAGRLVREABPYVSSMAGLAQLPLRSAAWATAKV 238
Y 185 MVEEGIDPQAFADPOPTKGIRRVTERLNRADRLYRAATGVRLPLFPDRCPGIMAAGKI 244
Y 239 YRKIGVKEQAGKQAWDHQSTSTAETKLTLLLTASGOAVTSRMKTYPPR 287
Y 245 YAAIGAFAVAKYKYNITERRHTTKGRKLWLVANSAMSTATSMPLSPR 293

RESULT 6
PSY_NARPS
ID PSY_NARPS STANDARD; PRT; 423 AA.
AC P53797;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE GN Phytoene synthase, chloroplast precursor (EC 2.5.1.-).
OS Psy.
OS Narcissus pseudonarcissus (Daffodil).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;
OC Narcissus.
OX NCBI_TaxID=39639;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Paracolla;
RA Schledz M., Beyer P.;
RT "Nucleotide sequence of a Narcissus pseudonarcissus cDNA for phytoene
RT synthase.";
RL (in) Plant Gene Register PGR95-122.
CC -1- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
CC phytoene.
CC -1- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
CC prephytoene diphosphate.
CC -1- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
CC phytoene.
CC -1- PATHWAY: Carotenoid biosynthesis.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78814; CAA55391.1; -.
CC PIR; S54135; S54135.
CC InterPro; IPR002060; Sm/phyt_synthse.
CC InterPro; IPR008949; Terpenoid_synth.
CC Pfam; PF00494; SQS_PSI; 1.
CC PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
CC PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
CC Multifunctional enzyme; Carotenoid biosynthesis;
CC Isoprene biosynthesis; Transferase; Chloroplast; Transient peptide.
CC TRANSIT 137 423 CHLOROPLAST (POTENTIAL).
CC CHAIN 137 423 PHYTOENE SYNTHASE.
CC SEQUENCE 423 AA; 47805 MW; 8E76BA35312B73DF CRC64;

Query Match 21.4%; Score 328.5; DB 1; Length 423;
Best Local Similarity 27.9%; Pred. No. 7.8e-22;
Matches 81; Conservative 58; Mismatches 132; Indels 19; Gaps 5;

Y 5 SKSFATATLFDKTRRSVLMVYAWCRHCDVDVDDQTLGPHADQPSQMPQRLQOLEM 64
Y 140 AKTYLGLTLMTERRRAIWAIVWCERTDELVDGHN-----ASHITPSALDRWEAR 191
Y 65 TRQAYAGSQMHPEPAPAFQEVAMAH--DIAPAYAFPHLGFAMDVRETRYLTLDLRYC 122
Y 192 LEDLFAG-RPYDNFDALSDTVSRFPYDIQF--FMDWVEGRMDLKKSKYKNFDELYLYC 248
Y 123 YHVGAVGVGLMMAQIMGV-----RDNATLDRCADGLAQTLNIAARDIVDDAQVGRCYLP 176
Y 249 YVAGTVGLSVVPWNGIAPESLAEASVYNAALAGIANQLTNILDEDVGEDARRGRIYLP 308
Y 177 ESWLEEBGLTKANYAAPENRQALSRIAGRLVREABPYVSSMAGLAQLPLRSAAWATAK 236
Y 309 QDELAAGLSDEDVFTGKVTDKWRSFMKRIKARTFFEQAEKGVTELSQASRPWWASL 368
Y 237 QVYRKIGVKEQAGKQAWDHQSTSTAETKLTLLLTASGOAVTSRMKTYPP 286
Y 369 LLYRQILDEIANDYNNFTKRAYVSVKVRALALPLAYGKSLILPLSLRPP 418

RESULT 7
PSY_CAPAN

```



373 LTVRRILDEITANDYNNFTKRAYSVPKLLIALPIAYAKSLVPESTRT 419

Db

RESULT 8

PSY2\_LYCES

ID PSY2\_LYCES STANDARD; PRT; 310 AA.

AC P37273;

DT 01-OCT-1994 (Rel. 30, Created)

DT 28-OCT-1994 (Rel. 30, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phytoene synthase 2, chloroplast precursor (BC 2.5.1.-) (Fragment).

GN PSY2.

OS Lycopersicon esculentum (Tomato).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

OC lamiales; Solanales; Solanaceae; Solanum.

NCBI\_TaxID=4081;

RN [1]

RN SEQUENCE FROM N.A.

RP

RC TISSUE=Leaf;

RC MEDLINE=94064645; PubMed=8245008;

EX Bartley G.E., Scolnik P.A.;

FA "cDNA cloning, expression during development, and genome mapping of

RT PSY2, a second tomato gene encoding phytoene synthase.";

RL J. Biol. Chem. 268:25718-25721(1993).

CC -I- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to

CC phytoene.

CC -I- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +

CC prephytoene diphosphate.

CC -I- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +

CC phytoene.

CC -I- PATHWAY: Carotenoid biosynthesis.

CC -I- SUBUNIT: Monomer (By similarity).

CC -I- SUBCELLULAR LOCATION: Chloroplast.

CC -I- DEVELOPMENTAL STAGE: In mature leaves.

CC -I- SIMILARITY: Belongs to the phytoene/squalene synthetase family.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC at the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: L23424; AAA34187.1; -.

DR PIR: A49558; A49558.

DR InterPro: IPR002050; Squ/phyt\_synth.

DR InterPro: IPR008949; Terpenoid\_synth.

DR Pfam: PF00494; SQS\_PSY; 1.

DR PROSITE: PS01044; SQUALEN\_PHYTOEN\_SYN\_1; 1.

DR PROSITE: PS01045; SQUALEN\_PHYTOEN\_SYN\_2; 1.

KW Multifunctional enzyme; Carotenoid biosynthesis; Multigene family;

KW Isoprene biosynthesis; Transferase; Chloroplast; Transit peptide.

FT NON\_TER 1 1

FT TRANSIT <1 25 CHLOROPLAST (POTENTIAL).

FT CHAIN 26 310 PHYTOENE SYNTHASE 2.

SQ SEQUENCE 310 AA; 35224 MW; 8EAF60FED9F4CF5 CRC64;

Query Match 20.0%; Score 307.5; DB 1; Length 310;

Best Local Similarity 28.9%; Pred. No. 4e-20;

Matches 84; Conservative 53; Mismatches 129; Indels 25; Gaps 7;

Qy 5 SKSFATASITLFDKATRSVLMLYAWCHGDDVIDDQTLGFHADQPSQMPQELQOLEMK 64

Db 28 AKTFFYLGTLMLTPDRERATWAIYVWCRRITDELVDGN-----ASHITPQALDREAR 79

Qy 65 TRQAYAGSQHSPAPFAPOEVAMAH--DIAPAYAFHLEGFAMDVRETRYLTLDLTLRYC 122

Db 80 LEDIFNG-RPFQMLDRLALSDTVSRPVPDIQPR--DMVSGMRDLKSYRNFEYLILYC 136

Qy 123 YHVAGYVGLMAAQIMGV--RDNATLD-----RACDLGLAFQALFNIAIDI VDDAQVGCYLP 176

b 137 YVAGTVGLMSVPMGIAPEKATTSVYVNAALAGIANQLNLRLDVGEDARRGRVYLP 196  
 Y 177 ESWLEEGTLCANYAAPENRQALSRIAGSLVREAPYVSSMAGLAQLPLRSAMAIATAK 236  
 b 197 ODELAQGLSDEIDIPAGKVTDKWIFPKWQIQARKEFFDEAEKGVTELSASASWPVLASL 256  
 Y 237 QVTRKIGVKVQKQAGKQANDHROSTSTAEKLTLLLTASGOAVTSRMKTYPPR 287  
 b 257 LLYRKILDEIENDYNNFTTRAYVSPKXLLTLPTIAYAFSLV-----PPK 301

## RESULT 9

SY DAUCA STANDARD; PRT; 398 AA.  
 C QSSUS;  
 T 16-OCT-2001 (Rel. 40, Created)  
 T 16-OCT-2001 (Rel. 40, Last sequence update)  
 T 28-FEB-2003 (Rel. 41, Last annotation update)  
 E Phytoene synthase, chloroplast precursor (EC 2.5.1.-).  
 N PSY.  
 S Daucus carota (Carrot).  
 C Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 C Campanulids; Apiales; Apiaceae; Apiaceae; Scandiceae; Daucinae;  
 C Daucus.  
 X NCBI\_TaxID=4039;  
 N [1]  
 P SEQUENCE FROM N.A.  
 C TISSUE=ROOT;  
 A Uno T., Sankawa U.;  
 T "Daucus carota phytoene synthase.";  
 L Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 L FUNCTION: Catalyzes the reaction from prephytoene diphosphate to  
 C phytoene.  
 C -!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +  
 C prephytoene diphosphate.  
 C -!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +  
 C phytoene.  
 C -!- PATHWAY: Carotenoid biosynthesis.  
 C -!- SUBUNIT: Monomer (By similarity).  
 C -!- SUBCELLULAR LOCATION: Chloroplast.  
 C -!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.

-----  
 C This SWISS-PROT entry is copyright. It is produced through a collaboration  
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 C the European Bioinformatics Institute. There are no restrictions on its  
 C use by non-profit institutions as long as its content is in no way  
 C modified and this statement is not removed. Usage by and for commercial  
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 C or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 C -----

EMBL; AB032797; BA84763.1; .  
 InterPro; IPR02060; Squ/phyt synthase.  
 InterPro; IPR008949; Terpenoid synth.  
 Pfam; PF00494; SQS\_PSY\_1  
 PROSITE; PS01044; SQUALEN\_PHTOEN SYN 1; 1.  
 PROSITE; PS01045; SQUALEN\_PHTOEN SYN 2; 1.  
 C Multifunctional enzyme; Carotenoid biosynthesis;  
 C Isoprene biosynthesis; Transferase; Chloroplast; Transit peptide.  
 FT TRANSIT 1 CHOROPLAST (POTENTIAL).  
 CHAIN ? 398 PHYTOENE SYNTHASE.  
 SEQUENCE 398 AA; 45199 MW; 1D1E043824730615 CRC64;

Query Match 20.0%; Score 307.5; DB 1; Length 398;  
 Best Local Similarity 25.5%; Pred. No. 5.4e-20;  
 Matches 85; Conservative 57; Mismatches 113; Indels 33; Gaps 9;

5 SKSFATATLFDKTRSLVLMYAWCRCHCDVIDDQTLGPHADQPSQMPQRLQLEMK 64  
 122 AKTYLGLTLPTRPRAVAIYVWCRTSLVDGPN-----ASHITPKALDRWEKR 173  
 65 TRQVAGSQMHEPAPAEQEVAMAH--DIAPAVAFDHEGFAMDVRETRYLTLDOTLRVC 122

Db 174 LNDLFDG-QPYDMYDAALADTYSTVPDIQPK--DMIDGEMDLKKSRYQTFDESLYLC 230  
 Qy 123 YFVAGVGLMAQIMGV--RDNATLD----RACDLGLAFQLTNIARDIVDDAQGRCYLP 176  
 Db 231 YVAGTVGLMSVPMGIAPEKATTSVYVNAALAGIANQLNLRLDVGEDARRGRVYLP 290  
 Qy 177 ESWLEEGTLCANYAAPE-----NRQALSRIAGSLVREAPYVSSMAGLAQLPLRSA 229  
 Db 291 QEBELKLAGIT-----PEVIFKGVTKWRSFKMQQ--IKRARMFFDEAEKGVTELSASAS 343  
 Qy 230 WAIATAKQVYRKIGVKVQKQAGKQANDHROSTSTAEKLTLLLTASGOAV 277  
 Db 344 MPVWASILLYKQILDIAEANDYNTFTKRAYVGKAKKLVSPLAYSRAAL 391

## RESULT 10

PSY ARATH STANDARD; PRT; 422 AA.  
 AC P37271; Q22375; Q8LEB6;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Phytoene synthase, chloroplast precursor (EC 2.5.1.-).  
 GN PSY1 OR PSY OR ATSG17230 OR MKP11.8.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 CX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=94286755; PubMed=8016277;  
 RA Scolnik P.A., Bartley G.B.;  
 RT "Nucleotide sequence of an Arabidopsis cDNA for phytoene synthase.";  
 RL plant Physiol. 104:1471-1472(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Castagnano P., Giuliano G.;  
 RT "Sequence of the phytoene synthase gene of Arabidopsis.";  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=97471969; PubMed=9330910;  
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
 RA Miyajima N., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
 RT features of the 1.6 Mb regions covered by twenty physically assigned  
 RT P1 clones";  
 RL DNA Res. 4:215-230(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 RA Feldmann K.A.;  
 RT "Full-length cDNA from Arabidopsis thaliana";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=22954850; PubMed=14593172;  
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
 RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
 RA Chankawa T., Ban J., Banno P., Bowser L., Brooks S.Y., Carninci P.,  
 RA Arakawa T., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,  
 RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
 RA Kaniya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,

Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.:  
"Empirical analysis of transcriptional activity in the Arabidopsis  
genome.";  
Science 302:842-846(2003).

-!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to  
phytoene.

-!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +  
prephytoene diphosphate.

-!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +  
phytoene.

-!- PATHWAY: Carotenoid biosynthesis.

-!- SUBUNIT: Monomer (By similarity).

-!- SUBCELLULAR LOCATION: Chloroplast.

-!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L25812; AAA32836.1; ..

EMBL; AF009954; AAB65697.1; ..

EMBL; AB005238; BAB10510.1; ..

EMBL; AY085565; AAM62787.1; ..

EMBL; BT000450; AAN17427.1; ..

EMBL; BT002084; AAN72095.1; ..

InterPro; IPR002060; Squ/phyt\_synthase.

InterPro; IPR008949; Terpenoid\_synth.

Pfam; PF00494; SOS\_PSY; 1.

PROSITE; PS01044; SQUALEN\_PHYTOEN SYN 1; 1.

PROSITE; PS01045; SQUALEN\_PHYTOEN SYN 2; 1.

Multifunctional enzyme; Carotenoid biosynthesis;

Isoprene biosynthesis; Transferrase; Chloroplast; Transit peptide.

TRANSIT 1 70 CHLOROPLAST (POTENTIAL).

CHAIN 71 422 PHYTOENE SYNTHASE.

CONFLICT 60 60 R -> M (IN REF. 4).

CONFLICT 128 128 L -> LV (IN REF. 1).

CONFLICT 143 143 A -> P (IN REF. 1).

SEQUENCE 422 AA; 47486 MW; C44FOA512F2DD31E CRC64;

Query Match 20.0%; Score 307; DB 1; Length 422;

Best Local Similarity 27.8%; Pred. No. 6.5e-20;

Matches 81; Conservative 58; Mismatches 130; Indels 22; Gaps 7;

5 SKSFATSTLFDKTRRSVLMYAKRCHDDVIDDQTLGFHADQPSOMPEORLQLEMK 64

143 AKTFYGLTLMTPERRKAIWAIYVWCRTDSELVDGPN-----ASHITPMALDRWEAR 194

65 TRQAVAGSQMHEPAPAFQEVAMAH--DIAPAYAFDHLGFMADVRETRYLTLDLRYC 122

195 LEDLFRG-RPFDMLDAALADTVARYEPVDIQPPR--DMIEGRMDLKKSRQNFDDLYLC 251

123 YHVAGVGLMAQIMGV--RDNATLD---RACDGLAFQLTNIARDIVDDAQVGRCYLP 176

252 YVAGTVGLMSVPVWGIDPKSKATTESVYNAALAGIANQLTNILRDVGEDARRGRVLP 311

177 ESWLEEGTUKANYAAPENRQALSRIAGLRVREAPFYVSSWAGLAQLPLRSAMAIATAK 236

312 QBELAQAGLSDEIDFAGKVTQWRNFMKQKRAEMFFDEAEKGVTELSAASRPFWASL 371

237 QVYRKIGVKVQAGKQAWDRHQSTSTAEKLTLLTLLTASQAVTSRMKTVPPR 287

372 LLYRQILDEIANDYNNFTKRAYGVKGKLLALPLAYAKSV---LNTSSSR 419

RESULT 11

SY MAIZE

D\_PSY\_MAIZE

C\_P49085;

01-FEB-1996 (Rel. 33, Created)

01-FEB-1996 (Rel. 33, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)

Phytoene synthase, chloroplast precursor (EC 2.5.1.1-).

Y1.

Zea mays (Maize).

Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC PCCAD clade; Panicoideae; Andropogoneae; Zea.

NCBI\_TaxID=4577;

(1)

SEQUENCE FROM N.A.

MEDLINE=96304610; PubMed=8722797;

Buckner B., Sammguel P., Janick-Buckner D., Bennetzen J.L.;

"The Y1 Gene of maize codes for phytoene synthase.";

Genetics 143:479-488(1996).

-!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to

phytoene.

-!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +

prephytoene diphosphate.

-!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +

phytoene.

-!- PATHWAY: Carotenoid biosynthesis.

-!- SUBUNIT: Monomer (By similarity).

-!- SUBCELLULAR LOCATION: Chloroplast.

-!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U32636; AAB60314.1; ..

PIR; S68307; S68307.

MaizeDB; 66643; ..

InterPro; IPR002060; Squ/phyt\_synthase.

InterPro; IPR008949; Terpenoid\_synth.

Pfam; PF00494; SOS\_PSY; 1.

PROSITE; PS01044; SQUALEN\_PHYTOEN SYN 1; 1.

PROSITE; PS01045; SQUALEN\_PHYTOEN SYN 2; 1.

Multifunctional enzyme; Carotenoid biosynthesis;

Isoprene biosynthesis; Transferrase; Chloroplast; Transit peptide.

TRANSIT 1 410 CHLOROPLAST (POTENTIAL).

CHAIN 344 344 N -> T (IN ALLELE B73).

VARIANT 344 344 N -> T (IN ALLELE B73).

SEQUENCE 410 AA; 46481 MW; 21070A33624ED79 CRC64;

Query Match 19.7%; Score 302.5; DB 1; Length 410;

Best Local Similarity 28.1%; Pred. No. 1.6e-19;

Matches 79; Conservative 53; Mismatches 130; Indels 19; Gaps 6;

5 SKSFATSTLFDKTRRSVLMYAKRCHDDVIDDQTLGFHADQPSOMPEORLQLEMK 64

130 AKTFYGLTLMTPERRKAIWAIYVWCRTDSELVDGPNY-----ITPTALDRWEKR 181

65 TRQAVAGSQMHEPAPAFQEVAMAH--DIAPAYAFDHLGFMADVRETRYLTLDLRYC 122

182 LEDLFTG-RPYDMLDAALSDTISRFPIDIQPPR--DMIEGRMDLKKRYNFDLYNYC 238

123 YHVAGVGLMAQIMGV--RDNATLD---RACDGLAFQLTNIARDIVDDAQVGRCYLP 176

239 YVAGTVGLMSVPVWGIDPKSKATTESVYNAALAGIANQLTNILRDVGEDARRGRVLP 298

177 ESWLEEGTUKANYAAPENRQALSRIAGLRVREAPFYVSSWAGLAQLPLRSAMAIATAK 236

299 QBELAQAGLSDEIDFAGKVTQWRNFMKQKRAEMFFDEAEKGVTELSAASRPFWASL 358

237 QVYRKIGVKVQAGKQAWDRHQSTSTAEKLTLLTLLTASQAV 277

359 LLYRQILDEIANDYNNFTKRAYGVKGKLLALPLAYAKSL 399



```

-!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; L27652; AA19428.1; -.
PIR; A53583; A53583.
InterPro; IPR002060; Squ/phyt synthetase.
InterPro; IPR008949; Terpenoid synth.
Pfam; PF00494; SQS_PSY; 1.
PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.
PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
Multifunctional enzyme; Carotenoid biosynthesis; Transmembrane;
Isoprene biosynthesis; Transferase.
TRANSMEM 6 26 POTENTIAL.
TRANSMEM 30 50 POTENTIAL.
TRANSMEM 76 96 POTENTIAL.
TRANSMEM 118 138 POTENTIAL.
TRANSMEM 146 166 POTENTIAL.
TRANSMEM 168 188 POTENTIAL.
TRANSMEM 230 250 POTENTIAL.
SEQUENCE 602 AA; 58927 MW; DDFBBD725007B5C2 CRC64;

Query Match 19.3%; Score 297; DB 1; Length 602;
Best Local Similarity 28.6%; Pred. No. 7.9e-19;
Matches 88; Conservative 57; Mismatches 105; Indels 58; Gaps 10;

Y 5 SKGFATASTLFDKATRSVLMYANCRHCDVDVDD-----QTLGFPA-----DQ 48
b 295 SRSYLLANSFLSGRLIDILLYSLFCLADLDVDAKSRREVLSTAKLNHFLDLHYKDA 354
Y 49 PSSQMEQRLQLEMTKTRQAYAGSOMHEPAPAFQEVAM--AHDIAPAYAFDHLGFMAD 106
b 355 DATEDPKKAERIDAYIKTAPP-----CAYQALHLLPHLPKPLDYDLKGFEMD 406
Y 107 VRETRYLTLD-----DTLYCYHVAGVVG-LMMAQI-----MGVRDNATLD 146
b 407 SQTFHGTSDSLDQVFIADDDXDLNLYAVAGTVGELCIAIIVHCLPDMSDTKQKRE 466
Y 147 -RACDGLAFLNTIARDIVDDAOCRCYLPSWLEEEGLT-KANYAAPENKQALSRIAG 204
b 467 TAACRGIALQVYNIARDIVDARIGRVLPITWLKXGLTHKVLNBPGEPIERNR 526
Y 205 RLVEAEPPYVSSMAGLAQLPLRSAAWAIATAKQVYRKIG-----VKVQAGKQAW 254
b 527 RLLENAPELYGGARPQMQRIPSEARGPMIGAVENTYMAIGRVLRERKEGTVFVRMEGRATV 586
Y 255 DHRQSTST 262
b 587 PKRRRLST 594

RESULT 14
SY_CUCMBE STANDARD; PRT; 422 AA.
D -PSY_CUCMBE
T C P49293;
T T 01-FEB-1996 (Rel. 33, Created)
T T 01-FEB-1996 (Rel. 33, Last sequence update)
T T 28-FEB-2003 (Rel. 41, Last annotation update)
E Phytoene synthase, chloroplast precursor (EC 2.5.1.-) (MEL5).
N PSY.
S Cucumis melo (Muskmelon).
S Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
C Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
C eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.
X NCBI_TaxID=3656;
N [1]
P SEQUENCE FROM N.A.

```

```

RC STRAIN=cv. Cantaloup Charentais; TISSUE=Pericarp;
RX MEDLINE=95284364; PubMed=7766896;
RA Karvouni Z., John I., Taylor J.E., Watson C.F., Turner A.J.,
RA Grierson D.;
RT isolation and characterisation of a melon cDNA clone encoding
RT phytoene synthase";
RL Plant Mol. Biol. 27:1153-1162(1995).
CC -!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
CC phytoene.
CC -!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
CC prephytoene diphosphate.
CC -!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
CC phytoene.
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z37543; CA895775.1; -.
DR PIR; S56688; S56688.
DR InterPro; IPR002060; Squ/phyt synthetase.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00494; SQS_PSY; 1.
DR PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.
DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
KW Multifunctional enzyme; Carotenoid biosynthesis;
KW Isoprene biosynthesis; Transferase; Chloroplast; Transit peptide.
FT TRANSIT 1 83
FT CHAIN 84 422 PHYTOENE SYNTHASE.
SQ SEQUENCE 422 AA; 47392 MW; 259124B3AD4642B0 CRC64;

Query Match 18.6%; Score 286; DB 1; Length 422;
Best Local Similarity 26.5%; Pred. No. 4.9e-18;
Matches 76; Conservative 56; Mismatches 125; Indels 30; Gaps 6;

QY 5 SKGFATASTLFDKATRSVLMYANCRHCDVDVDDQTLGFHADQSSQMEQRLQLEMK 64
Db 148 AKTFYLTGTLMTPEQKAIWAIYVWCRTDDELVDGPN-----ASHITPTALDRWEAR 199
QY 65 TRQAYAGSQMHEPAPAFQEVAMAH-----DIAPAYAFDHLGFMADVRETRYLTLD 117
Db 200 LEEFQGRFPD-----MLDALADTVTKFPVDIQPK--DMIEGHEMDLRKSRYKNFDE 251
QY 118 TLRYCYHVAGVVGLMVAQINGV-----RDNATLDRACDLGLAFL-TNIAARDIVDDAQV 170
Db 252 LDYCYVYVAGTVGLMSVPVNGIAPESQASTESVYNAALALGIANQAPPNILRDVGEDARR 311
QY 171 GRCYLPSWLEEEGLTKANYAAPENKQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAAW 230
Db 312 GRILYLPDELAQAGLSDEDFAGRVTDKWRNFKNCKIKARMFFDEAEKGVLELNKASRW 371
QY 231 AIATAKQVYRKIGVKVEQAGKQAWDHRQSTSTAEKLTLLLTASQAV 277
Db 372 PVWASULLYRQILDEIEANDYDNFTKRAYVSKAKKILALPMAYGRAL 418

RESULT 15
CRTY SPIPL STANDARD; PRT; 309 AA.
ID -CRTY SPIPL
AC O07333;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene synthase (EC 2.5.1.-).
GN CRTB OR PYS.

```

```

NS Spirulina platensis.
X Bacteria; Cyanobacteria; Oscillatoriales; Spirulina.
X NCBI_TaxID=1156;
UN [1]
IP SEQUENCE FROM N.A.
IP STRAIN=IAM M-135;
A Kato Y., Yano S., Kojima H.;
L Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
L -!- FUNCTION: Catalyzes the reaction from prephytoene diphosphate to
C phytoene.
C -!- CATALYTIC ACTIVITY: 2 geranylgeranyl diphosphate = diphosphate +
C prephytoene diphosphate.
C -!- CATALYTIC ACTIVITY: Prephytoene diphosphate = diphosphate +
C phytoene.
C -!- PATHWAY: Carotenoid biosynthesis.
C -!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
C -----
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation -
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (see http://www.isb-sib.ch/announce/
C or send an email to license@isb-sib.ch).
C -----
R EMBL; AB001284; BAA20384.1; -.
R InterPro; IPR002060; Squ/phyt_synthase.
R InterPro; IPR008949; Terpenoid_synth.
R Pfam; PF00494; SQS_PSY; 1.
R PROSITE; PS01044; SQUALEN_PHYTOEN_SYN_1; 1.
R PROSITE; PS01045; SQUALEN_PHYTOEN_SYN_2; 1.
W Multifunctional enzyme, Carotenoid biosynthesis; Transferase.
Q SEQUENCE 305 AA; 35345 MW; C42A1C6431604C75 CRC64;
-----
Query Match 18.2%; Score 280; DB 1; Length 309;
Best Local Similarity 23.7%; Pred. No. 1.1e-17;
Matches 86; Conservative 44; Mismatches 128; Indels 32; Gaps 8;

2Y 5 SKSFATASTLPDAKTRSVLMYAWCHDCDDVIDDQTLGPHADQPSSQMPQRLQQLMK 64
DB 31 AKTYLGLQMLPLAKQAIWAIYVWCRTBELVDGPM-----ASSTLTLDHWEHQ 82

2Y 65 TRQYAGSQMHEPAAFOEV--AMAHDIAPAYAFDHLGFMVDRETRYLTLDOTLYC 122
DB 83 LESIFAGHPI-EPVDVALVDTLGRFPLDIQPER--DVIAGQMDLSRNYNTPELNLIC 139

2Y 123 YHVGIVGLMAQIMGV-----RDNATL---DRACDLGLAQLTNIARDIVDDAQ 169
DB 140 YRVAGTVGLMSLAYNGTAEPDLSVPWNRDQSIYYPKEEALGLAQLTNILDVGEDAR 199

2Y 170 VRCVLPESWLEEEGLTKA---NYAAPENQALSRIAGRLVREAEPPYVSSMAGLAQLPL 226
DB 200 RGRILPLDLDLALFNFTYEDLLNGKVDWRWELMRFO---IQARKEYTLAEEGIAALHP 256

2Y 227 RSANAIAATKQVYKIGVKVQKQAWDHQSTSTAEKLTLLLTASQQA 276
DB 257 DIRNPVWLTALMYRQILDEIERNEYDFNQRAYVPTWKMMCLPLAQLRA 306

```

Search completed: February 29, 2004, 14:45:14  
Job time : 6.98876 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

n on: February 29, 2004, 14:34:14 ; Search time 8.23902 Seconds  
(without alignments)  
3455.835 Million cell updates/sec

tle: US-09-941-947A-34

rfect score: 1535

quence: 1 MAVGSKSFATSTLFDANKR.....VTSRMKTYPPRAHLWQRP 296

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

tal number of hits satisfying chosen parameters: 283366

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : PIR 78:\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1371	89.3	296	2 S52587	prephytoene pyroph
2	1368	89.1	296	2 E37802	crb protein - Erw
3	1028	67.0	309	2 B39273	geranylgeranyl-dip
4	395.5	25.8	344	2 T30895	prephytoene pyroph
5	394	25.7	355	2 S49621	phytoene synthase
6	393.5	25.6	339	2 S0403	geranylgeranyl-dip
7	385	25.1	355	2 T50746	phytoene synthase
8	362	23.6	325	2 D75466	phytoene synthase
9	328.5	21.4	423	2 S54135	phytoene synthase
10	307.5	20.0	310	2 A49558	phytoene synthase
11	302.5	19.7	410	2 S68307	phytoene synthase
12	300.5	19.6	412	2 A42102	phytoene synthase
13	300	19.5	336	2 S32170	geranylgeranyl-dip
14	297	19.3	602	2 A53583	phytoene synthase
15	286	18.6	412	2 S22474	geranylgeranyl-dip
16	286	18.6	422	2 S56668	phytoene synthase
17	276.5	18.0	425	2 T10702	phytoene synthase
18	266	17.3	310	2 A2035	phytoene synthase
19	265.5	17.3	337	2 S45360	phytoene synthase
20	261.5	17.0	290	2 A81118	phytoene synthase
21	258	16.8	307	2 S20383	probable poly-isop
22	255.5	16.6	290	2 H81902	probable phytoene
23	248.5	16.0	331	2 T36969	probable phytoene
24	228.5	14.9	312	2 T35400	probable phytoene
25	224.5	14.6	319	2 T46594	probable phytoene
26	217.5	14.2	302	2 B70735	probable phytoene
27	214.5	14.0	322	2 H84299	phytoene synthase
28	210.5	13.7	299	2 F69108	phytoene synthase
29	209.5	13.6	322	2 E84320	phytoene synthase

30	201.5	13.1	277	2 G90469	phytoene synthase
31	196.5	12.8	275	2 H81074	probable transfe
32	193.5	12.6	276	2 D81868	probable transfe
33	171.5	11.2	417	2 F52090	squalene synthase
34	171.5	11.2	417	2 A45998	farnesyl-diphospha
35	171.5	11.2	417	2 T38245	farnesyl-diphospha
36	165	10.7	362	2 T45141	farnesyl-diphospha
37	164.5	10.7	287	2 C90061	squalene desaturas
38	163.5	10.7	416	2 S52075	farnesyl-diphospha
39	163	10.6	390	2 A84226	farnesyl-diphospha
40	159.5	10.4	416	2 A45105	farnesyl-diphospha
41	144.5	9.4	303	2 T35399	probable phytoene
42	140.5	9.2	274	2 G69837	phytoene synthase
43	135	8.8	299	2 T51118	phytoene synthase
44	134	8.7	254	2 A55548	crbM protein - Sta
45	132.5	8.6	460	2 B48057	farnesyl-diphospha

## ALIGNMENTS

## RESULT 1

S52587

prephytoene pyrophosphate synthase - Erwinia herbicola

C:Species: Erwinia herbicola

C:Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 02-Mar-2001

R:Accession: S52587

R:Lin. Y.P.: Lai, E.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.

Mol. Gen. Genet. 245, 417-423, 1994

A:Title: Transcriptional activation of flanking sequences by Tn1000 insertion.

A:Reference number: S52583; MUID:95107237; PMID:7808390

A:Accession: S52587

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-296 <LIN>

A:Cross-references: EMBL:M90698; NID:G148393; PIDN:AAA21264.1; PID:G148398

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1992

C:Superfamily: Mycobacterium marinum phytoene synthase

Query Match 89.3%; Score 1371; DB 2; Length 296;

Best Local Similarity 88.5%; Pred. No. 7e-116;

Matches 262; Conservative 12; Mismatches 22; Indels 0; Gaps 0;

QY	1	MAVGSKSFATSTLFDANKRSLVLMYACRCHDDVDDQTLGFHADQPSQMEQRIQQ	60
DB	1	MAVGSKSFATSKLFDANKRSLVLMYACRCHDDVDDQTLGFHADQPSQMEQRIQQ	60
QY	61	LEMKTRQAYAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGEPAMDVRETRYLFLD	120
DB	61	LEMKTRQAYAGSQMHEPAPAFQEVAMAHDIAPAYAFDHLGEPAMDVRETRYLFLD	120
QY	121	YCHVAGVGLMMAQINGVRDNATLDRACDGLAFQLTNIARDIVDDAQCVCYL	180
DB	121	YCHVAGVGLMMAQINGVRDNATLDRACDGLAFQLTNIARDIVDDAQCVCYL	180
QY	181	EEEGTKANYAPENRQALSRIAGRLVREAPYVSSMAGLAQLPLRSAMAIATAKQYR	240
DB	181	EEEGTKANYAPENRQALSRIAGRLVREAPYVSSMAGLAQLPLRSAMAIATAKQYR	240
QY	241	KIGKVQAGKQAWDHROSTSTAEKLTLLLTASQAVTSRMKTYPPRAHLWQRP	296
DB	241	KIGKVQAGKQAWDHROSTSTAEKLTLLLTASQAVTSRMKTYPPRAHLWQRP	296

## RESULT 2

E37802

crb protein - Erwinia uredovora

C:Species: Erwinia uredovora

C:Date: 31-May-1991 #sequence\_revision 31-May-1991 #text\_change 02-Mar-2001

C:Accession: E37802

R:Misawa, N.; Nakagawa, M.; Kobayashi, K.; Yamano, S.; Izawa, Y.; Nakamura, K.; Harashim

J. Bacteriol. 172, 6704-6712, 1990

A:Title: Elucidation of the Erwinia uredovora carotenoid biosynthetic pathway by functio

;Reference number: A37802; MUID:91072214; PMID:2254247

;Accession: B37802

;Status: preliminary

;Molecule type: DNA

;Residues: 1-296 <MIS>

;Cross-references: GB:D90087; NID:g216681; PIDN:BAAL4128.1; PID:d1014831; PID:g216686

;Superfamily: Mycobacterium marinum phytoene synthase

Query Match 89.1%; Score 1368; DB 2; Length 296;

Best Local Similarity 88.9%; Pred. No. 1.3e-115;

Matches 263; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

Y 1 MAVGSKSPATASTLFDKATRSVLMYAWCRHCDVDIDDTLGFHADQPSQPEQLQQ 60

b 1 MAVGSKSPATASTLFDKATRSVLMYAWCRHCDVDIDDTLGFHADQPSQPEQLQQ 60

Y 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120

b 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120

Y 121 YCHVAGVGLMAQIMGVDRNATIDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

b 121 YCHVAGVGLMAQIMGVDRNATIDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

Y 181 EEEGLTKANYAAPENRQALSRAGLVREAEFYVSSMAGLAQLPLRSAMAIATAKQVYR 240

b 181 EEEGLTKANYAAPENRQALSRAGLVREAEFYVSSMAGLAQLPLRSAMAIATAKQVYR 240

Y 241 KIGVKVQAGKQANDHROSTSTAEKLTLLTASGOAVTSRMKTYPPRPAHLWQRP 296

b 241 KIGVKVQAGKQANDHROSTSTAEKLTLLTASGOAVTSRMKTYPPRPAHLWQRP 296

RESULT 3

339273

teranylgeranyl-diphosphate geranylgeranyltransferase (BC 2.5.1.32) - Erwinia herbicola

;Alternate names: prephytoene pyrophosphate synthase

;Species: Erwinia herbicola

;Date: 16-Sep-1992 #sequence revision 16-Sep-1992 #text\_change 02-Mar-2001

;Accession: B39273; B3120

;Armstrong, G.A.; Alberti, M.; Hearst, J.E.

Proc. Natl. Acad. Sci. U.S.A. 87, 9975-9979, 1990

;Title: Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nor

;Reference number: A39273; MUID:91088634; PMID:2263648

;Accession: B39273

;Molecule type: DNA

;Residues: 1-309 <ARM>

;Cross-references: GB:X38423; NID:g148401; PIDN:AAA24821.1; PID:g148403

;Gene: CrtB

;Superfamily: Mycobacterium marinum phytoene synthase

;Keywords: transferase

Query Match 67.0%; Score 1028; DB 2; Length 309;

Best Local Similarity 64.9%; Pred. No. 6.2e-85;

Matches 192; Conservative 36; Mismatches 68; Indels 0; Gaps 0;

Y 1 MAVGSKSPATASTLFDKATRSVLMYAWCRHCDVDIDDTLGFHADQPSQPEQLQQ 60

b 14 MANGSKSPATASTLFDKATRSVLMYAWCRHCDVDIDDTLGFHADQPSQPEQLQQ 73

Y 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120

b 74 LRTLTAAEFGAEMQDPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 133

Y 121 YCHVAGVGLMAQIMGVDRNATIDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

b 134 YCHVAGVGLMAQIMGVDRNATIDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 193

Y 181 EEEGLTKANYAAPENRQALSRAGLVREAEFYVSSMAGLAQLPLRSAMAIATAKQVYR 240

b 194 QDGLTPENYAAENRQALSRAGLVREAEFYVSSMAGLAQLPLRSAMAIATAKQVYR 253

QY 241 KIGVKVQAGKQANDHROSTSTAEKLTLLTASGOAVTSRMKTYPPRPAHLWQRP 296

DB 254 EIGIKVRAAGSANDRQHTSKGKIAMLMARAPQGVIRAKTTFTVPRPAGLWQRPV 309

RESULT 4

T50895

prephytoene pyrophosphate synthase [imported] - Rubrivivax gelatinosus

;Species: Rubrivivax gelatinosus

;Date: 21-Jul-2000 #sequence revision 21-Jul-2000 #text\_change 02-Mar-2001

;Accession: T50895

;Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.

submitted to the EMBL Data Library, November 1999

;Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosyn

;Reference number: Z525270

;Accession: T50895

;Status: preliminary; translated from GB/EMBL/DDBJ

;Molecule type: DNA

;Residues: 1-344 <NAG>

;Cross-references: EMBL:AB034704; PIDN:BA94048.1

;Experimental source: strain IL144

;Genetics:

;Gene: crtB

;Superfamily: Mycobacterium marinum phytoene synthase

Query Match 25.8%; Score 395.5; DB 2; Length 344;

Best Local Similarity 36.2%; Pred. No. 7.3e-28;

Matches 106; Conservative 32; Mismatches 140; Indels 15; Gaps 3;

QY 1 MAVGSKSPATASTLFDKATRSVLMYAWCRHCDVDIDDTLGFHADQPSQPEQLQQ 60

DB 18 MRGSKSFFAASLLLPQVTRTPATATYAFRVADDAV-----LSGDPHAAWDE 66

QY 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGEGFAMDVRETRYLTLDLTLR 120

DB 67 LRTLDVAYAGTAPTAADRALASTVHRVGVPRVLLDALLLEGFLWDAGRRYDTIADVEA 126

QY 121 YCHVAGVGLMAQIMGVDRNATIDRACDGLAFQLTNIARDIVDDAQVGRCYLPESWL 180

DB 127 YGARVAGTGAAGAAALMGVRSPOALARACELGVAMQFTNIARDVGDGADARGLYLPREWL 186

QY 181 EEEGLTKANYAAPENRQALSRAGLVREAEFYVSSMAGLAQLPLRSAMAIATAKQVY 239

DB 187 VEAGLDVDAVLQNPVHCEVAQVTRVRLRAADBELYSEHSHGIAALPDCRPAIRAARLVY 246

QY 240 KIGVKVQAGKQANDHROSTSTAEKLTLLTASGOAVTSRMKTY---PPRPA 289

DB 247 AEIGRMLEKGLDSVNRVVPVPPARRKALMARASAAFTNTPGRATISMPPLEPA 299

RESULT 5

S49621

phytoene synthase - Rhodobacter sphaeroides

;Species: Rhodobacter sphaeroides

;Date: 05-Mar-1995 #sequence revision 12-May-1995 #text\_change 02-Mar-2001

;Accession: S49621

;Lang, H.P.; Cogdell, R.J.; Takaichi, S.; Hunter, C.N.

submitted to the EMBL Data Library, November 1994

;Description: The complete DNA sequence, specific TMS insertion map and gene assignment

;Reference number: S49619

;Accession: S49621

;Status: preliminary

;Molecule type: DNA

;Residues: 1-355 <LAN>

;Cross-references: EMBL:X82458; NID:g575405; PID:g575408

;Genetics:

;Gene: crtB

;Superfamily: Mycobacterium marinum phytoene synthase

Query Match 25.7%; Score 394; DB 2; Length 355;

Best Local Similarity 36.8%; Pred. No. 1e-27;

Matches 105; Conservative 32; Mismatches 128; Indels 22; Gaps 3;









Search completed: February 29, 2004, 14:52:54  
Job time : 10.239 secs

RESULT 15  
 22474.  
 phytoene synthase (EC 2.5.1.-) - tomato  
 ;Alternate names: ripening protein 5  
 ;Species: Lycopersicon esculentum (tomato)  
 ;Date: 19-Mar-1997 #sequence revision 18-Jul-1997 #text change 02-Mar-2001  
 ;Accession: S22474; S06321; S35155; S35154; S21981; S24968  
 ;Ray, J.; Moureau, P.; Bird, C.; Bird, A.; Grierson, D.; Maunders, M.; Truesdale, M.; E  
 ;Title: Cloning and characterization of a gene involved in phytoene synthesis from toma  
 ;Reference number: S22474; MUID:92322971; PMID:1623189  
 ;Accession: S22474  
 ;Molecule type: DNA  
 ;Residues: 1-412 <RAY>  
 ;Cross-references: EMBL:X60441; GB:S39572; NID:g19248; PIDN:CAA42969.1; PID:g19249  
 ;Note: the authors did not translate the codon for residue 155  
 ;Ray, J.; Bird, C.; Maunders, M.; Grierson, D.; Schuch, W.  
 ;Title: Sequence of pTOM5, a ripening related cDNA from tomato.  
 ;Reference number: S06321; MUID:88096591; PMID:3697097  
 ;Accession: S06321  
 ;Status: not compared with conceptual translation  
 ;Molecule type: mRNA  
 ;Residues: 1-406,'N',408-412 <RAY1>  
 ;Cross-references: GB:Y00521; NID:g19340; PIDN:CAA68575.1; PID:g19341  
 ;Fray, R.G.; Grierson, D.  
 ;Title: Identification and genetic analysis of normal and mutant phytoene synthase gene  
 ;Reference number: S35154; MUID:93344508; PMID:8343597  
 ;Accession: S35155  
 ;Status: translation not shown  
 ;Molecule type: mRNA  
 ;Residues: 1-387,'NMLKDFNFNFKSRKGSNATITVLGLAPCE','TYIADRDGPIGITF' <PRA1>  
 ;Cross-references: EMBL:X67144; NID:g19346; PIDN:CAA47625.1; PID:g19347  
 ;Experimental source: mutant ry  
 ;Accession: S35154  
 ;Status: translation not shown  
 ;Molecule type: mRNA  
 ;Residues: 1-107,'VECLSDPLQTERGTGVSLPQGVGSGAT','RWGPGGRRRGADGTQCCIGLIFC' <PRA2>  
 ;Cross-references: EMBL:X67143; NID:g19398; PIDN:CAA47624.1; PID:g19399  
 ;Experimental source: mutant r  
 ;Genetics:  
 ;Gene: Psyl  
 ;Introns: 138/1; 155/1; 212/3; 291/2; 355/3  
 ;Superfamily: Mycobacterium marinum phytoene synthase  
 ;Keywords: chloroplast; transferase  
 Query Match 18.6%; Score 286; DB 2; Length 412;  
 Best Local Similarity 29.7%; Pred. No. 6.9e-18;  
 Matches 79; Conservative 46; Mismatches 119; Indels 22; Gaps 6;  
 y 5 SKGATATLFDKTRSRVLMYAKCRHDDVIDQTLGFHADPSSQMPQRLOQLEMK 64  
 b 132 AKTNLTGMLTPTERRRATWAIYVWCRRDVELDQPN-----ASYITPAALDRWNR 183  
 y 65 TRQAYAGSQMHEPAFAFOEVA-MAHDIAPAYAFDHELEGFAMRVRETRYLLDITLRYCY 123  
 b 184 LEDVFNCRPPDMLDGALSDTVSNFFVDIQPER--DMIEGWRDLRKSRYKNFDLYLYCY 241  
 y 124 HVAGVGLGMAAQMGV--RDNATLD-----RACDLGLAFQLTNIARDIVDDAQVGRVILPE 177  
 b 242 YVAGTVGLMGVPTMGIAPEASKATTESVYNAALALGIANQLTNILRDVGEDARRGRVLPQ 301  
 y 178 SWLEEGELTKYAAAPENRQALSRIAGRUVRAEPTVYSSMAGLAQLPLRSAMAIATAKQ 237  
 b 302 DELAQAGLSDEDIFAGRVTDTKWRFPMKKQIHARKEFFDEAEKGVTELSASRFPFWASLV 361  
 y 238 VYRKIGVYKVEQAGQKQWDRHQSTSTA 263  
 b 362 LYRKILDEIE-----ANDYNNFTKRA 382

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

n on: February 29, 2004, 14:51:24 ; Search time 19.6527 Seconds  
(without alignments)  
3180.293 Million cell updates/sec

tle: US-09-941-947A-34

fect score: 1535  
quence: 1 MAVGSKSFATASTLFDKTR.....VTSRMKTYPRFAHLWQRP 296

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 809742 seqs, 211153259 residues

tal number of hits satisfying chosen parameters: 809742

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubaa/US09C\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	ID	Description
1	1535	100.0	296	10	US-09-941-947A-34
2	1535	100.0	296	14	Sequence 34, Appl
3	686.5	44.7	303	9	US-10-218-118-10
4	686.5	44.7	303	10	Sequence 10, Appl
5	367.5	23.9	225	16	US-09-920-923-3
6	367.5	23.9	225	16	Sequence 3, Appl
7	310.5	20.2	410	9	US-10-038-854-133
8	310.5	20.2	410	9	Sequence 133, App
9	302.5	19.6	412	9	US-09-847-081B-2
10	300.5	19.6	412	9	Sequence 2, Appl
11	300.5	19.6	412	15	US-10-038-854-130
12	270.5	17.6	402	15	Sequence 130, App
13	232	15.1	342	14	US-10-038-854-131
14	229.5	15.0	316	14	US-10-038-854-131
15	213	13.9	314	14	Sequence 131, App
					Sequence 76, Appl
					Sequence 318, Appl
					Sequence 8563, App
					Sequence 9192, App
					Sequence 16, Appl

16	201.5	13.1	1268	15	US-10-438-784-3	Sequence 3, Appl
17	196.5	12.8	275	15	US-10-438-784-7	Sequence 7, Appl
18	178	11.6	292	9	US-09-738-626-6173	Sequence 6173, Ap
19	172	11.2	374	9	US-09-820-004-2	Sequence 2, Appl
20	171.5	11.2	417	9	US-09-820-004-4	Sequence 4, Appl
21	171.5	11.2	417	9	US-09-820-004-5	Sequence 5, Appl
22	171.5	11.2	417	9	US-09-820-004-6	Sequence 6, Appl
23	168.5	11.0	363	14	US-10-358-917-2	Sequence 2, Appl
24	164.5	10.7	287	14	US-10-358-917-10	Sequence 10, Appl
25	164	10.7	279	16	US-10-038-854-16	Sequence 16, Appl
26	163	10.6	390	15	US-10-369-493-18574	Sequence 18574, A
27	159.5	10.4	416	14	US-10-205-194-67	Sequence 67, Appl
28	155.5	10.1	255	16	US-10-038-854-128	Sequence 128, App
29	143.5	9.3	520	14	US-10-128-714-3505	Sequence 3505, Ap
30	143.5	9.3	528	14	US-10-128-714-8505	Sequence 8505, Ap
31	141.5	9.2	354	15	US-10-369-493-13039	Sequence 13039, A
32	140	9.1	304	9	US-09-738-626-4195	Sequence 4195, Ap
33	139.5	9.1	448	14	US-10-032-585-7355	Sequence 7355, Ap
34	134	8.7	254	16	US-10-038-854-129	Sequence 129, App
35	131.5	8.6	300	14	US-10-156-761-9193	Sequence 9193, Ap
36	130.5	8.5	460	15	US-10-369-493-2381	Sequence 2381, Ap
37	126.5	8.2	413	15	US-10-429-949-5	Sequence 5, Appl
38	124	8.1	444	15	US-10-369-493-22063	Sequence 22063, A
39	121.5	7.9	356	15	US-10-369-493-3321	Sequence 3321, Ap
40	118	7.7	194	16	US-10-038-854-356	Sequence 356, App
41	116	7.6	388	14	US-10-024-130A-6	Sequence 6, Appl
42	116	7.6	410	14	US-10-024-130A-2	Sequence 2, Appl
43	111	7.2	38	16	US-10-038-854-132	Sequence 132, App
44	111	7.2	328	14	US-10-156-761-11454	Sequence 11454, A
45	109	7.1	131	14	US-10-149-759-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1

US-09-941-947A-34  
; Sequence 34, Application US/09941947A  
; Publication No: US20030003528A1  
; GENERAL INFORMATION:  
; APPLICANT: Brzostowicz, Patricia C.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: DiCosimo, Deana J.  
; APPLICANT: Koffas, Mattheos  
; APPLICANT: Miller, Edward S. Jr.  
; APPLICANT: Odom, J. Martin  
; APPLICANT: Picatoggio, Steve  
; APPLICANT: Rouviere, Pierre E.  
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE  
; FILE REFERENCE: C11903 US NA  
; CURRENT APPLICATION NUMBER: US/09/941,947A  
; CURRENT FILING DATE: 2001-09-01  
; PRIOR APPLICATION NUMBER: 60/229,907  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,858  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 34  
; LENGTH: 296  
; TYPE: PRT  
; ORGANISM: Pantoea stewartii  
US-09-941-947A-34

Query Match	100.0%;	Score	1535;	DB 10;	Length	296;	
Best Local Similarity	100.0%;	Pred. No.	2.7e-160;				
Matches	296;	Conservative	0;	Mismatches	0;	Gaps	0;
Indels							
Qy	1	MAVGSKSFATASTLFDKTRRSVLMYAWCRHCDVDVDDDTLGFHADQPSQPEQRLOQ	60				
Db	1	MAVGSKSFATASTLFDKTRRSVLMYAWCRHCDVDVDDDTLGFHADQPSQPEQRLOQ	60				
Qy	61	LEMKTRQYAGSQMHSPAFAAQFQEVAMAHDIAPAYADHLEGFAMDVRETRYLTLDITLR	120				

b 61 LEMKTRQAVAGSOMHEPFAAFQEVAMAHDIAPAFDELEGFAMDVRETRYLTLDLTLR 120  
y 121 YCVHVAGVYGLMMAQIMGVDRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180  
b 121 YCVHVAGVYGLMMAQIMGVDRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180  
y 181 EBEGLTKANYAAPENROALSRIAGRLVREABPYVYSSMAGLAQLPLRSAMAIATAKQVYR 240  
b 181 EBEGLTKANYAAPENROALSRIAGRLVREABPYVYSSMAGLAQLPLRSAMAIATAKQVYR 240  
y 241 KIGVKVEQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296  
b 241 KIGVKVEQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296

RESULT 2

S-10-218-118-10  
Sequence 10, Application US/10218118  
Publication No. US20030148319A1  
GENERAL INFORMATION:  
APPLICANT: Brzostowicz, Patricia  
APPLICANT: Rouviere, Pierre  
APPLICANT: Picaataggio, Stephen  
APPLICANT: Cheng, Qiong  
TITLE OF INVENTION: Genes Encoding Carotenoid Compounds  
FILE REFERENCE: CL1876 US NA  
CURRENT APPLICATION NUMBER: US/10/218,118  
CURRENT FILING DATE: 2002-08-13  
PRIOR APPLICATION NUMBER: 60/312,646  
PRIOR FILING DATE: 2001-08-15  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 10  
LENGTH: 296  
TYPE: PRT  
ORGANISM: Pantoea stewartii

S-10-218-118-10

Query Match 100.0%; Score 1535; DB 14; Length 296;  
Best Local Similarity 100.0%; Pred. No. 2.7e-160;  
Matches 296; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 1 MAVGSKSFATSTLFDKTRRSVLMYAWCRCHDDVIDDQTLGFHADQPSSQMPERLQQ 60  
b 1 MAVGSKSFATSTLFDKTRRSVLMYAWCRCHDDVIDDQTLGFHADQPSSQMPERLQQ 60  
y 61 LEMKTRQAVAGSOMHEPFAAFQEVAMAHDIAPAFDELEGFAMDVRETRYLTLDLTLR 120  
b 61 LEMKTRQAVAGSOMHEPFAAFQEVAMAHDIAPAFDELEGFAMDVRETRYLTLDLTLR 120  
y 121 YCVHVAGVYGLMMAQIMGVDRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180  
b 121 YCVHVAGVYGLMMAQIMGVDRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYLPESWL 180  
y 181 EBEGLTKANYAAPENROALSRIAGRLVREABPYVYSSMAGLAQLPLRSAMAIATAKQVYR 240  
b 181 EBEGLTKANYAAPENROALSRIAGRLVREABPYVYSSMAGLAQLPLRSAMAIATAKQVYR 240  
y 241 KIGVKVEQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296  
b 241 KIGVKVEQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296

RESULT 3

S-09-547-267-3  
Sequence 3, Application US/09547267  
Patent No. US20020147371A1  
GENERAL INFORMATION:  
APPLICANT: Hohmann, Hans-Peter  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tessier, Michel  
APPLICANT: van Loon, Adolphus

;; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
;; NUMBER OF SEQUENCES: 47  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Hoffmann-La Roche Inc.  
;; STREET: 340 Kingsland Street  
;; CITY: Nutley  
;; STATE: NJ  
;; COUNTRY: USA  
;; ZIP: 07110  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/547,267  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/660,645  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Pokras, Bruce A.  
;; REGISTRATION NUMBER: 32,748  
;; REFERENCE/DOCKET NUMBER: RAN 6002/170  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (201) 235-5801  
;; TELEFAX: (201) 235-2363  
;; INFORMATION FOR SEQ ID NO: 3:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 303 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-547-267-3

Query Match 44.7%; Score 686.5; DB 9; Length 303;

Best Local Similarity 48.7%; Pred. No. 7.8e-67;  
Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;

Qy 1 MAVGSKSFATSTLFDKTRRSVLMYAWCRCHDDVIDDQTLGFHADQPSSQMPERLQQ 60  
Db 12 IAQGSQSFQAQAKLPPGIREDTVMYAWCRHADDVIDGQVMSAPAEAGD--PQARLGA 69  
Qy 61 LEMKTRQAVAGSOMHE-----PAPAFQEVAMAHDIAPAFDELEGFAMDVRETRYLT 115  
Db 70 LRADTLAA-----LHEDGPMSPFFAALRCVARRHDFPDLWPMDLIEGFAMDVADREYKSL 124  
Qy 116 DDTLRYCYHVAGVYGLMMAQIMGVDRDNATLDRACDLGLAFQLTNIARDIVDDAQVGRCYL 175  
Db 125 DDVLEYSHVAGVYGVMMARVMGVQDDAVIDRACDLGLAFQLTNIARDIVDDAIGRCYL 184  
Qy 176 PESWLEEBGLTKANYAAPENROALSRIAGRLVREABPYVYSSMAGLAQLPLRSAMAIATA 235  
Db 185 PADMLAEAG---ATVEGPPVSDALYSVIRLLDAAEPTYSASARQGLPHLPCCAWSIAAA 241  
Qy 236 KQYRKIGVKVQAGKQAWDRHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 295  
Db 242 LRIYRAIGTRIQGGPEATYRQIRISTSKAKTOLLARGGLDAAASRLRGCEISRDLGWTPEP 301

RESULT 4

US-09-920-923-3  
Sequence 3, Application US/09920923  
Publication No. US2003002273A1  
GENERAL INFORMATION:  
APPLICANT: Pasamontes, Luis  
APPLICANT: Taygenkov, Yuri  
TITLE OF INVENTION: Improved Fermentative Carotenoid Production  
FILE REFERENCE: Improved Fermentative Carotenoid  
CURRENT APPLICATION NUMBER: US/09/920,923  
CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 08/980,832  
PRIOR FILING DATE: 1997-12-01  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 303  
TYPE: PRT  
ORGANISM: Flavobacterium sp. R1534  
-09-920-923-3  
Query Match 44.7%; Score 686.5; DB 10; Length 303;  
Best Local Similarity 48.7%; Pred. No. 7.8e-67;  
Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;  
1 MAVGSKSPAFATLTFDAKTRSVLMVWCHCHDDVDDDTGLGFHADQPSQMPERLQQ 60  
12 IAQSQSPFAQAKLMPGIREDTVNLVWCHHDDVDDQVMSAPFAGD--PQARLGA 69  
61 LEMTRQAYAGSOME-----PAPAFQEVAMAHDTAPAYAFDEHGFAMVRETRYLTL 115  
70 LRADTLAA-----LHEDGPMSPPPAALRQVARRHDFDLWPMOLIEGFAMDVADREYRSI 124  
116 DDTLYRYCHVAGVGLMAQIMGVDRNATLDRACDGLAFOLTNLARDIVDDAQVGRCYL 175  
125 DDVLEYSTHVAGVGVWMAKVMGVQDDAVDRACDGLAFOLTNLARDIVDDAAIGRCYL 184  
176 PESMLEEBGLTKANYAAPENKASRIAGLRVREAPYVSSMAGLAQLPLRSAMAIATA 235  
185 PADMLAEAG--ATVEGVPFDALYSVILRLDAAEPYASARQGLPHLPCCAWSIAAA 241  
236 KQVTRKIGVKVQKQAKQWHRQSTSTAEKLTLLLTASGOAVTSRMKTYPPRPAHLWQRP 295  
242 LRITYRAIGTRIQGGPEAYRQRTSTSKAAKIGLLARGGLDAAASRLRGCEISRDGLWTRP 301

SULT 5  
-10-038-854-133  
Sequence 133, Application US/10038854  
Publication No. US20040022781A1  
GENERAL INFORMATION:  
APPLICANT: Spytex, Kimberly A  
APPLICANT: Li, Li  
APPLICANT: Wolenc, Adam R  
APPLICANT: Vernet, Corine  
APPLICANT: Eisen, Andrew J  
APPLICANT: Liu, Xiaohong  
APPLICANT: Malyankar, Uziel M  
APPLICANT: Shimkets, Richard A  
APPLICANT: Tchernev, Velizar  
APPLICANT: Spaderna, Steven K  
APPLICANT: Gorman, Linda  
APPLICANT: Kekuda, Ramesh  
APPLICANT: Patturajan, Meera  
APPLICANT: Gusev, Vladimir Y  
APPLICANT: Gangolli, Esha A  
APPLICANT: Guo, Xiaofei S  
APPLICANT: Shenoy, Suresh G  
APPLICANT: Rastelli, Luca  
APPLICANT: Casman, Stacie J  
APPLICANT: Soldog, Ferenc  
APPLICANT: Burgess, Catherine E  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Ellerman, Karen  
APPLICANT: Gunther, Erik  
APPLICANT: Smithson, Glenna  
APPLICANT: Millet, Isabelle  
APPLICANT: MacDougall, John R  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-230  
CURRENT APPLICATION NUMBER: US/10/038,854  
CURRENT FILING DATE: 2003-01-22  
PRIOR APPLICATION NUMBER: 60/259,928  
PRIOR FILING DATE: 2000-12-29

PRIOR APPLICATION NUMBER: 60/259,415  
PRIOR FILING DATE: 2001-01-02  
PRIOR APPLICATION NUMBER: 60/259,785  
PRIOR FILING DATE: 2001-01-04  
PRIOR APPLICATION NUMBER: 60/269,814  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/279,832  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/279,833  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/279,863  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/283,889  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/284,447  
PRIOR FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 60/286,683  
PRIOR FILING DATE: 2001-04-25  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 411  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 133  
LENGTH: 225  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-038-854-133  
Query Match 23.9%; Score 367.5; DB 16; Length 225;  
Best Local Similarity 46.4%; Pred. No. 6.8e-32;  
Matches 85; Conservative 22; Mismatches 57; Indels 19; Gaps 5;  
QY 102 GFMDVR-----STRVLTDDTLRYCHVAGVGLMAQIMGVDR-----NATLD 146  
DB 43 GMAMDLEKREKNLQRYATFEDLLRYCYVAGTVGLMMARLGMVGRKLEDPADWOLEBVD 102  
QY 147 -RACDGLAFOLTNLARDIVDDAQVGR--YLPESWLEEBGLTKANYAAPENR-QALSRI 202  
DB 103 LRACDGLALQTLNARDVGEDARSGPCRVYLPTEWLSQYGLSLEDLLAPENTDKRIRRV 162  
QY 203 AGLRVRAEPYVSSMAGLAQLPLRSAMAIATAKQVTRKIGVKVQKQAKQWHRQSTST 262  
DB 163 LRRLDNRARAYEDALTGLAGLPQSRFPFIAAAPQVYAGIGDAIEANGYDVFRRAKTRK 222  
QY 263 AEK 265  
DB 223 GEK 225  
RESULT 6  
US-09-847-081B-2  
Sequence 2, Application US/09847081B  
Patent No. US20020128464A1  
GENERAL INFORMATION:  
APPLICANT: BAYER AG  
TITLE OF INVENTION: DNA encoding the tobacco phytoene synthase  
FILE REFERENCE: Le A 34 326  
CURRENT APPLICATION NUMBER: US/09/847,081B  
CURRENT FILING DATE: 2001-05-02  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 440  
TYPE: PRT  
ORGANISM: Nicotiana tabacum  
US-09-847-081B-2  
Query Match 21.9%; Score 336.5; DB 9; Length 440;  
Best Local Similarity 30.3%; Pred. No. 4.7e-28;  
Matches 88; Conservative 52; Mismatches 131; Indels 19; Gaps 6;  
QY 5 SKSPATASTLFDKATRSVLMVWCHCHDDVDDDTGLGFHADQPSQMPERLQQLEMK 64  
DB 158 AKTYVLGTKLTPERRAIWAIYVWCRTDSELVDGPN-----ASHITPQALDRWETR 209

Y 65 TRQAYAGSQHPEPAPAFQEVAMAH--DIAPAYAFDHLEGPAMDVRETRYLTLDLTLRYC 122  
b 210 LEDIFSG-RPFDMLDAALSDTVSRFPVDIQPR--DMIEGMEMLKSRKTFDELYLC 266  
Y 123 YHVAGVVGGLMAQINGV--RDNATLD----RACDLGLAFQLTNIARDIVDDAQVGRCYLP 176  
b 267 YTVAGTVGLNSVPVNGIAPESKATTESVYNALALGLANQLTNILRDVGEDARGRVYLP 326  
Y 177 ESWLEBEGLTKANYAPENRQALSRIAGRLVREABPYVSSMAGLAQPLRSANAIATAX 236  
b 327 QDELAQAGLSDEDDIFAGRVTDKRNFMKKQIORAKFPDESEKGVTELDLSASRPVLTAL 386  
Y 237 QVYRKIGVKGVEQAGQAMDHROSTSTAKLTLLTASQAVTSRMKTYPP 286  
b 387 LLYRKILDEIANDYNNFTFRAYVSKPKLLTLPAYAKSLVPPNRTSSP 436

## RESULT 7

S-09-847-081B-4

Sequence 4, Application US/09847081B

Patent No. US20020128464A1

GENERAL INFORMATION:

APPLICANT: BAYER AG

TITLE OF INVENTION: DNA encoding the tobacco phytoene synthase

FILE REFERENCE: Le A 34 326

CURRENT APPLICATION NUMBER: US/09/847,081B

CURRENT FILING DATE: 2001-05-02

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 4

LENGTH: 410

TYPE: PRT

ORGANISM: Nicotiana tabacum

FEATURE:

NAME/KEY: unsure

LOCATION: 135, 139

OTHER INFORMATION: Xaa is unknown or other

S-09-847-081B-4

Query Match 20.2%; Score 310.5; DB 9; Length 410;  
Best Local Similarity 28.2%; Pred. No. 3.1e-25;  
Matches 81; Conservative 56; Mismatches 131; Indels 19; Gaps 5;

Y 5 SKSFATASTLFDKATRRSVMLYAWCRHCDVDIDQTLGFHADQPSQMPQRLQOLEMK 64  
b 130 AKTFYGTXTMTPEERRAIWAIYVWCERTDELVDGN-----ASHITPQALDRWESR 181  
Y 65 TRQAYAGSQHPEPAPAFQEVAMAH--DIAPAYAFDHLEGPAMDVRETRYLTLDLTLRYC 122  
b 182 LEDVFSG-RPFDMLDAALSDTVSKFPVDIQPR--DMIEGMEMLKSRKTFDELYLC 238  
Y 123 YHVAGVVGGLMAQINGV--RDNATLD----RACDLGLAFQLTNIARDIVDDAQVGRCYLP 176  
b 239 YTVAGTVGLNSVPVNGIAPESKATTESVYNALALGLANQLTNILRDVGEDARGRVYLP 298  
Y 177 ESWLEBEGLTKANYAPENRQALSRIAGRLVREABPYVSSMAGLAQPLRSANAIATAX 236  
b 299 QDELAQAGLFDDEDDIFAGRVTDKRNFMKKQIORAKFPDESEKGVTELDLSASRPVLTAL 358  
Y 237 QVYRKIGVKGVEQAGQAMDHROSTSTAKLTLLTASQAVTSRMKTYPP 283  
b 359 LLYRKILDEIANDYNNFTFRAYVSKPKLLTLPAYAKSLVPPNRTSSP 405

## RESULT 8

S-10-038-854-130

Sequence 130, Application US/10038854

Publication No. US20040022781A1

GENERAL INFORMATION:

APPLICANT: spytek, Kimberly A

APPLICANT: Li, Li

APPLICANT: Wolenc, Adam R

S-10-038-854-130

APPLICANT: Vernet, Corine  
APPLICANT: Bisen, Andrew J  
APPLICANT: Liu, Xiaohong  
APPLICANT: Malyankar, Uriel M  
APPLICANT: Shinkov, Richard A  
APPLICANT: Tchernev, Velizar  
APPLICANT: Spaderna, Steven K  
APPLICANT: Gorman, Linda  
APPLICANT: Kekuda, Ramesh  
APPLICANT: Patturajan, Meera  
APPLICANT: Gusev, Vladimir Y  
APPLICANT: Gangolli, Bsha A  
APPLICANT: Guo, Xiaojia S  
APPLICANT: Shenoy, Suresh G  
APPLICANT: Rastelli, Luca  
APPLICANT: Casman, Stacie J  
APPLICANT: Boldog, Ferenc  
APPLICANT: Burgess, Catherine E  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Ellerman, Karen  
APPLICANT: Gunther, Erik  
APPLICANT: Smithson, Glenda  
APPLICANT: Millet, Isabelle  
APPLICANT: MacDougall, John R  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-230  
CURRENT APPLICATION NUMBER: US/10/038,854  
CURRENT FILING DATE: 2003-01-22  
PRIOR APPLICATION NUMBER: 60/258,928  
PRIOR FILING DATE: 2000-12-29  
PRIOR APPLICATION NUMBER: 60/259,415  
PRIOR FILING DATE: 2001-01-02  
PRIOR APPLICATION NUMBER: 60/259,785  
PRIOR FILING DATE: 2001-01-04  
PRIOR APPLICATION NUMBER: 60/269,814  
PRIOR FILING DATE: 2001-02-20  
PRIOR APPLICATION NUMBER: 60/279,832  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/279,833  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/279,863  
PRIOR FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/283,889  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: 60/284,447  
PRIOR FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 60/286,683  
PRIOR FILING DATE: 2001-04-25  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 411  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 130  
LENGTH: 436  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-10-038-854-130

Query Match 20.2%; Score 310.5; DB 16; Length 436;  
Best Local Similarity 27.8%; Pred. No. 3.4e-25;  
Matches 80; Conservative 58; Mismatches 131; Indels 19; Gaps 5;

QY 5 SKSFATASTLFDKATRRSVMLYAWCRHCDVDIDQTLGFHADQPSQMPQRLQOLEMK 64  
b 156 AKTFYGTXTMTPEERRAIWAIYVWCERTDELVDGN-----ASHITPQALDRWESR 207  
QY 65 TRQAYAGSQHPEPAPAFQEVAMAH--DIAPAYAFDHLEGPAMDVRETRYLTLDLTLRYC 122  
b 208 LEDLFRG-RPFDMLDAALSDTVTKFPVDIQPR--DMIEGMEMLKSRKTFDELYLC 264  
QY 123 YHVAGVVGGLMAQINGV--RDNATLD----RACDLGLAFQLTNIARDIVDDAQVGRCYLP 176  
b 265 YTVAGTVGLNSVPVNGIAPESKATTESVYNALALGLANQLTNILRDVGEDARGRVYLP 324





362 LYRKILDEIENDYNNFTKRAYVSKKLIPIAYAKSLVPPTKT 407

3-10-401-321-76  
Sequence 76, Application US/10401321  
Publication No. US20030233679A1  
GENERAL INFORMATION:

APPLICANT: Brown, Sherri M.  
APPLICANT: Heck, Gregory R.  
APPLICANT: Pillier, Kenneth J.  
APPLICANT: Kilshore, Ganesh M.  
APPLICANT: Ellich, Ted D.  
APPLICANT: Logusch, Eugene W.  
APPLICANT: Rao, Sudabathula  
APPLICANT: Ream, Joel E.  
APPLICANT: Logusch, Sherry J.  
APPLICANT: Baerson, Scott R.

TITLE OF INVENTION: Methods for Controlling Gibberellin Levels

FILE REFERENCE: 11899.0216.DVUS01 (MOBT:216--1)

CURRENT APPLICATION NUMBER: US/10/401,321

CURRENT FILING DATE: 2003-03-27

NUMBER OF SEQ ID NOS: 89

SOFTWARE: PatentIn version 3.2

SEQ ID NO 76

LENGTH: 412

TYPE: PRT

ORGANISM: Lycopersicon esculentum

S-10-401-321-76

Query Match 19.6%; Score 300.5; DB 15; Length 412;  
Best Local Similarity 29.0%; Pred. No. 4e-24;  
Matches 83; Conservative 51; Mismatches 135; Indels 17; Gaps 5;

5 SKSFATASTLFDKATRSVLMYVWCRHDDVDQTLGFHADQSSQMPQRLQ 64

132 AKTFNLGTMLTPERRAIWAIYVWCRDDELVDGPN-----ASYITPAALORWNR 183

65 TRQVAGSQMHPEAPAFQVMAHDIAFAYAFDHLGPFAMDVRETRYLTLDL 123

184 LEDVFNRPFTMDLGSALSDTVSNFPVDIQPFR--DMIEGNEXDLRSRYKNFDELYLYCY 241

124 HVAGWGLMAQINGV--RDNATLD-----RACDLGLAFQLTNIARDIVDDAQVGRCLPE 177

242 YVAGTVGLMSVPIINGIAPESKATTESVYNAALALGIANQLTNILRDVGEDARRGRVYLPQ 301

178 SWLEEGHGTANYAAPENROALSRIAGRLVEAEPPYVSSMAGLAQLPLRSAMAIATAQ 237

302 DELAQAGLSDSDIFAGRVTDKWRIFPMKKQIHRARKFDEAEKGVTELSASRPVWASLV 361

238 VYRKIGVKEQAGKQADHRSQSTSTAELKTLTLLTASQAVTSRMKT 283

362 LYRKILDEIENDYNNFTKRAYVSKKLIPIAYAKSLVPPTKT 407

RESULT 12

US-10-259-194A-318

Sequence 318, Application US/10259194A

Publication No. US20040010815A1

GENERAL INFORMATION:

APPLICANT: Lange, Markus B.  
APPLICANT: Ghassemian, Majid  
APPLICANT: Briggs, Steven P.  
APPLICANT: Cooper, Bret  
APPLICANT: Glazebrook, Jane  
APPLICANT: Goff, Stephen A.  
APPLICANT: Katagiri, Fumiyaki  
APPLICANT: Kreps, Joel  
APPLICANT: Moughamer, Todd  
APPLICANT: Provart, Nicholas  
APPLICANT: Ricke, Darrell  
APPLICANT: Zhu, Tong

TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES

FILE REFERENCE: 70029-NP

CURRENT APPLICATION NUMBER: US/10/259,194A

CURRENT FILING DATE: 2003-01-07

PRIOR APPLICATION NUMBER: US 60/325,277

PRIOR FILING DATE: 2001-09-26

PRIOR APPLICATION NUMBER: US 60/370,743

PRIOR FILING DATE: 2002-04-04

PRIOR APPLICATION NUMBER: US 60/370,620

PRIOR FILING DATE: 2002-04-04

NUMBER OF SEQ ID NOS: 662

SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta

SEQ ID NO 318

LENGTH: 402

TYPE: PRT

ORGANISM: Oryza sativa

FEATURE:

NAME/KEY: X-region

LOCATION: (58)..(58)

OTHER INFORMATION: Xaa = any naturally occurring amino acid

US-10-259-194A-318

Query Match 17.6%; Score 270.5; DB 15; Length 402;  
Best Local Similarity 27.2%; Pred. No. 7.7e-21;  
Matches 78; Conservative 60; Mismatches 126; Indels 23; Gaps 8;

5 SKSFATASTLFDKATRSVLMY----AWCRHDDVDQTLGFHADQSSQMPQRLQ 60

126 AKTFYLETQLMTPERKAVWAIYGMVLWCRDDELVDGPN-----SSYITPKALDR 177

61 LEMKTRQVAGSQMHPEAPAFQVMAHDIAFAYAF-DHLGPFAMDVRETRYLTLDL 119

178 WEKRLDLPFG-RPYDMYDAALSDTVSKFPVDIQPFR--DMIEGNEXDLRSRYKNFDELY 236

120 RYCYHVAGVYVGLMAQINGV--RDNATLD-----ATLDRACDLGLAFQLTNIARDIVDDAQVGR 173

237 LYCYVAGTVGLMTVPWGIAPDSKASTESVYNAALALGIANQLTNILRDVGEEER-GRI 295

174 YLPESWLEEGHGTANYAAPENROALSRIAGRLVEAEPPYVSSMAGLAQLPLRSAMAI 232

296 YLPDLDELAEGAEEDIFRGKVTWKWKMGQILR-ARLFFDEAEKGVHLDLSASRPV 354

233 ATAKQVYRKIGVKEQAGKQADHRSQSTSTAELKTLTLLTASQAVTS 279

355 LASLWLYRQILDALEANDYNNFTKRAYVWNAKKKLLSLPVAYARAAYA 401

RESULT 13

US-10-156-761-8563

Sequence 8563, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 8563

LENGTH: 342

TYPE: PRT

ORGANISM: Streptomyces avermitilis

US-10-156-761-8563

Query Match 15.1%; Score 232; DB 14; Length 342;  
Best Local Similarity 27.6%; Pred. No. 1.1e-16;  
Matches 90; Conservative 40; Mismatches 136; Indels 60; Gaps 12;  
6 KSPATATLFDKTRRSVLMYAWCRHCDVID--DQTLGFHADQPSQMPQOR---LOO 60  
32 KTYFLATRLLPVERPAVHALYGFARWADDIVSLDTTVG-----PAVRSALGR 81  
61 LEMKTRQAYAGSQMHEPAPFAAFQSVAMAHDIAPAYAFDHLEGFAMDVRETRYTLTLDITLR 120  
82 LQESLRTGLRGHSTFVVLALASTARVYADHGFSGDFMTAMRSDLVGTGYETADLRG 141  
121 YCYHVAGVVGGLMAQIMGV--RDNATLDRACDLGLAFQLTNIARDIVDDAQVRCYLPES 177  
142 YMHGSAVIGLQMLPVLVGTGVVPREAA-PRHAAALGVAFQLTNFIROVGEDLDGRVILPA 200  
178 SWLEEEGLTK--ANYAAPENR-----QALSRIAG---RLVRAEAPYVSSMAGLAQLPL 226  
201 DLLRAHGVDRLLRSDTGRDRRIITRALKAVEGLTRGVYREAA-----GLAWLDP 253  
227 RSAWAIATAKOVYRKIGVKEQAGKQAWDHROSTSTASK-----LTLTLASQAV 277  
254 VARPCTRTAFVLYGGILDADDDG-YAVVHRRVAVPRRRAAVALDGLVRLGVARLGSRA 312  
278 TSMKTYTP-----PREAH 290  
313 DARLPTLPDWRPLALVPRSSPREAH 338

SULT 14  
-10-156-761-9192  
Sequence 9192, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: HOSIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156.761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 9192  
LENGTH: 316  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
-10-156-761-9192

Query Match 15.0%; Score 229.5; DB 14; Length 316;  
Best Local Similarity 30.0%; Pred. No. 1.8e-16;  
Matches 81; Conservative 38; Mismatches 134; Indels 17; Gaps 9;  
5 SKSPATATLFDKTRRSVLMYAWCRHCDVIDDQTLGFHADQPSQMPQORLQOLEMK 64  
30 ARNFAYGIRLLPTPKRRAMSAVAFPSRVDIDGQ---ALAPDVKAARLEDTTR--ALLSR 84  
65 TRQAYAGSQMHEPAPFAAFQSVAMAHDIAPAYAFDHL-EGFAMDVRETRYTLTLDITLR 123  
85 VRDGRVDEDDTPDPAVALAHAAEQFPI-PLGGLDELIDGLMDVGRGETYETWDDLVYCR 143  
124 HVAGVVGGLMAQIMGVRONAT-LRACD-----IGLAFQLTNIARDIVDDAQVRCYLPES 178  
144 CVAGAIGRVSLGVTFGTFCGARGARASAYADTLGLALQLTNILEDVREDAEGRKTYLPAD 203

QY 179 WLEEGLTANYAAPENRQALSRIAGRL---VRBAEPPYVSSMAGLAQLPLRSNAIATA 235  
DB 204 DLAKFGCS-AGFDRPIPEG-SDPAGLVHFEVRARALFAEGYRLLPMLDRRSACVAAK 261  
QY 236 KOVYRKIGVKEQAGKQAWDHROSTSTASK 265  
DB 262 AGIYRLLDRIEREPEAVLRGVSLPGREX 291

RESULT 15  
US-10-128-713A-16  
; Sequence 16, Application US/10128713A  
; Publication No. US20030170847A1  
; GENERAL INFORMATION:  
; APPLICANT: Bramucci, Michael G  
; TITLE OF INVENTION: Genes involved in Isoprenoid Compound Production  
; FILE REFERENCE: CL-1788  
; CURRENT APPLICATION NUMBER: US/10/128.713A  
; CURRENT FILING DATE: 2002-04-22  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16  
; LENGTH: 314  
; TYPE: PRT  
; ORGANISM: Rhodococcus erythropolis  
US-10-128-713A-16

Query Match 13.9%; Score 213; DB 14; Length 314;  
Best Local Similarity 27.7%; Pred. No. 1.1e-14;  
Matches 83; Conservative 46; Mismatches 143; Indels 28; Gaps 8;  
QY 6 KSPATATLFDKTRRSVLMYAWCRHCDVIDDQTLGFHADQPSQMPQORLQOLEMK 65  
DB 20 RTYFLATRLLPERRRAVHALYAFARVVDVDEPS-GPH-ERGTVLADVERAAVTALDN 77  
QY 66 RQAYAGSQMHEP-----AFAAQEVAMAHDIAPAYAFDHLEGFAMDVRETR----YLTL 115  
DB 78 PTATGGPFSTIPLDLTRVLPFAFADAVKTFDIPRAYDPAFFESMENDAPDTAKFRPVNTM 137  
QY 116 DDTLRYCYHVAGVVGGLMAQIMGV--DNATLDRACDLGLAFQLTNIARDIVDDAQVRC 173  
DB 138 DELAEYMYGSAVVGIGLQMLPILGVSVFQEAUVFASNLGEAFQLTNFIROVGEDLDGRGL 197  
QY 174 YLPSWLEEEGLTKANYAAPENRQALSRIAGRLVREAPY-----YVSSMAGLAQLPL 226  
DB 198 YLPAGEFAAFGV--DIEMLEHGRTGTGVIVRVKRALAHFIATVTRGYSAESGIPMLDR 254  
QY 227 RSAWAIATAKOVYRKIGVKEQAGKQAWDHROSTSTASKLTL-----LTLASQAVTSRMK 282  
DB 255 RVQPSIRTAFLVLYGAILDQVERADFRILHRRSVVPGTRLRVAAPGLVRSATYAANKRMR 314

Search completed: February 29, 2004, 15:28:05  
Job time : 22.6527 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

! protein - protein search, using sw model

in on: February 29, 2004, 14:35:44 ; Search time 8.99489 Seconds  
(without alignments)  
1698.885 Million cell updates/sec

tle: US-09-941-947A-34

fect score: 1535

quence: 1 MAVGSKSFATVASTLFDKTR.....VTGRMKTYPRPAHLWQRP1 296

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

tal number of hits satisfying chosen parameters: 389414

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

sult No.	Score	Query		Length	DB	ID	Description
		Match					
1	1368	89.1	296	1	US-07-783-705A-5		Sequence 5, Appli
2	1028	67.0	309	1	US-08-331-004A-2		Sequence 2, Appli
3	1028	67.0	309	5	PCT-US95-13937A-2		Sequence 2, Appli
4	997.5	65.0	308	1	US-08-095-726-6		Sequence 6, Appli
5	997.5	65.0	308	1	US-08-096-043-6		Sequence 6, Appli
6	997.5	65.0	308	1	US-08-093-577-6		Sequence 6, Appli
7	997.5	65.0	308	1	US-08-096-623A-6		Sequence 6, Appli
8	686.5	44.7	303	3	US-08-660-645A-3		Sequence 3, Appli
9	686.5	44.7	303	3	US-09-298-718-3		Sequence 3, Appli
10	686.5	44.7	303	3	US-09-546-969-3		Sequence 3, Appli
11	686.5	44.7	303	3	US-08-980-832-3		Sequence 3, Appli
12	686.5	44.7	303	4	US-09-547-267-3		Sequence 3, Appli
13	686.5	44.7	303	4	US-09-920-923B-3		Sequence 3, Appli
14	318.5	20.7	410	1	US-08-579-667-6		Sequence 6, Appli
15	318.5	20.6	410	1	US-08-579-667-2		Sequence 2, Appli
16	315.5	20.6	410	1	US-08-579-667-8		Sequence 8, Appli
17	311.5	20.3	413	1	US-08-579-667-4		Sequence 4, Appli
18	309.5	20.2	400	4	US-09-691-270A-14		Sequence 14, Appl
19	307.5	20.0	310	4	US-09-691-270A-27		Sequence 27, Appl
20	304.5	19.8	402	3	US-09-180-342-3		Sequence 3, Appli
21	302.5	19.7	410	4	US-09-691-270A-28		Sequence 28, Appl
22	286.5	18.7	408	4	US-09-691-270A-2		Sequence 2, Appli
23	252	16.4	252	4	US-09-691-270A-12		Sequence 12, Appl
24	249.5	16.3	248	4	US-09-691-270A-16		Sequence 16, Appl
25	243	15.8	242	4	US-09-691-270A-10		Sequence 10, Appl
26	227.5	14.8	186	4	US-09-691-270A-4		Sequence 4, Appli
27	179.5	11.7	673	4	US-09-091-725-13		Sequence 13, Appl

Sequence 19, Appli  
Sequence 2, Appli  
Sequence 6, Appli  
Sequence 4, Appli  
Sequence 5, Appli  
Sequence 6, Appli  
Sequence 23, Appli  
Sequence 7, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 9, Appli  
Sequence 8, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 78, Appli  
Sequence 4, Appli

## ALIGNMENTS

### RESULT 1

US-07-783-705A-5  
; Sequence 5, Application US/07783705A  
; Patent No. 5429939  
; GENERAL INFORMATION:  
; APPLICANT: Misawa, No. 5429939ihiko  
; APPLICANT: Kobayashi, Kazuo  
; APPLICANT: Nakamura, Katsumi  
; APPLICANT: Yamano, Shigeyuki  
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE  
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ladas & Parly  
; STREET: 26 West 61 Street  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10023  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: N/A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07783,705A  
; FILING DATE: 19911023  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 1-103078  
; FILING DATE: 21-APR-1989  
; APPLICATION NUMBER: JP 2-53225  
; FILING DATE: 05-MAR-1990  
; APPLICATION NUMBER: US 07/519,011  
; FILING DATE: 19-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Schwadron, Janet I.  
; REGISTRATION NUMBER: 33,778  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-708-1935  
; TELEFAX: 212-246-5959  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 296 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-783-705A-5

Query Match

89.1%; Score 1368; DB 1; Length 296;

Best Local Similarity 88.9%; Pred. No. 5.9e-143;  
Matches 263; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

Y 1 MAVGSKSPATSTLFDKATRRSVLMYAWCRHCDVDVDDQTLGPHADQPSQMPQRLOQ 60  
|||||  
b 1 MAVGSKSPATSTLFDKATRRSVLMYAWCRHCDVDVDDQTLGPHADQPSQMPQRLOQ 60  
|||||

Y 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYLTLDLTLR 120  
|||||  
b 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYLTLDLTLR 120  
|||||

Y 121 YCHVAGVGLMAQIMGVDRNATLDRACDLGLAQPLTNIAARDIVDDAQVGRCYLPESWL 180  
|||||  
b 121 YCHVAGVGLMAQIMGVDRNATLDRACDLGLAQPLTNIAARDIVDDAQVGRCYLPESWL 180  
|||||

Y 181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAAWAIATAKQVYR 240  
|||||  
b 181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAAWAIATAKQVYR 240  
|||||

Y 241 KIGVKVEQAGKQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296  
|||||  
b 241 KIGVKVEQAGKQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296  
|||||

## RESULT 2

S-08-331-004A-2

Sequence 2, Application US/08331004A

Patent No. 5618988

GENERAL INFORMATION:

APPLICANT: Hauptmann, Randal

APPLICANT: Eschenfeldt, William H

APPLICANT: English, Jami

APPLICANT: Brinkhaus, Friedhelm L

TITLE OF INVENTION: Enhanced Carotenoid Accumulation

TITLE OF INVENTION: in Storage Organs of Genetically

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amoco Corporation, Law Dept

STREET: 55 Shuman Boulevard, Suite 600

CITY: Naperville

STATE: IL

COUNTRY: USA

ZIP: 60563-8437

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/331,004A

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Galloway, No. 5618988val B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7087172447

TELEFAX: 7087172430

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 309 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

S-08-331-004A-2

Query Match 67.0%; Score 1028; DB 1; Length 309;

Best Local Similarity 64.9%; Pred. No. 2.7e-105;

Matches 192; Conservative 36; Mismatches 68; Indels 0; Gaps 0;

QY 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYLTLDLTLR 120  
|||||  
Db 74 LRTLTLLAAFEQEMQDPAPAFQEVALTHTGTPRMALDHLGFGFAMDVQATRYVTFTEDTLR 133  
|||||

QY 121 YCHVAGVGLMAQIMGVDRNATLDRACDLGLAQPLTNIAARDIVDDAQVGRCYLPESWL 180  
|||||  
Db 134 YCHVAGVGLMAQIMGVDRNATLDRACDLGLAQPLTNIAARDIVDDAQVGRCYLPESWL 193  
|||||

QY 181 EEEGLTKANYAAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAAWAIATAKQVYR 240  
|||||  
Db 194 QAGLTPENYAPENRQALSRIAGRLVREAEPPYVSSMAGLAQLPLRSAAWAIATAKQVYR 253  
|||||

QY 241 KIGVKVEQAGKQAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP 296  
|||||  
Db 254 EIGIKVKAAGGSANDRRQHTSKGKIAMLMAPGGVIRAKTTRVTRPAGLWQRPV 309  
|||||

## RESULT 3

PCT-US95-13937A-2

Sequence 2, Application PC/TUS9513937A

GENERAL INFORMATION:

APPLICANT: Hauptmann, Randal

APPLICANT: Eschenfeldt, William H

APPLICANT: English, Jami

APPLICANT: Brinkhaus, Friedhelm L

TITLE OF INVENTION: Enhanced Carotenoid Accumulation

TITLE OF INVENTION: in Storage Organs of Genetically

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amoco Corporation, Law Dept

STREET: 55 Shuman Boulevard, Suite 600

CITY: Naperville

STATE: IL

COUNTRY: USA

ZIP: 60563-8437

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/13937A

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Galloway, Norval B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 7087172447

TELEFAX: 7087172430

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 309 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US95-13937A-2

Query Match 67.0%; Score 1028; DB 5; Length 309;

Best Local Similarity 64.9%; Pred. No. 2.7e-105;

Matches 192; Conservative 36; Mismatches 68; Indels 0; Gaps 0;

QY 1 MAVGSKSPATSTLFDKATRRSVLMYAWCRHCDVDVDDQTLGPHADQPSQMPQRLOQ 60  
|||||  
Db 14 MANGSKSPATAAKLFDPATRRSVLMTWCRHCDVDVDDQTLGPHADQPSQMPQRLOQ 73  
|||||

QY 61 LEMKTRQAYAGSOMHEPAPAFQEVAMAHDIAPAYAFDHLGFGFAMDVRETRYLTLDLTLR 120  
|||||  
Db 74 LRTLTLLAAFEQEMQDPAPAFQEVALTHTGTPRMALDHLGFGFAMDVQATRYVTFTEDTLR 133  
|||||

QY 121 YCHVAGVGLMAQIMGVDRNATLDRACDLGLAQPLTNIAARDIVDDAQVGRCYLPESWL 180  
|||||

161	EEEGTITKANYAAPENRQALSRILAGLVRENEPYYSSMAGLAQLPLRSAWATATAKQVTR	240	
QY	194	QDAGLAPNTAARENRPALARWR-RLIDAAEPYIISQAGLHRLRRSAWATATARSVTR	252
DB	241	KIGVKVQAGKQAWDRHQSTSTAEKLTLLITAGQAVTSMKTYTPRPAHLHQRPPI	296
QY	253	EIGIKVKAAGSGAWDRHSTSKCKELAMLMAPPGVITRAKTRVTPRPAGLMORPV	308
DB			

## RESULTS

US-08-096-043-6  
 ? Sequence 6, Application US/08096043  
 ? Patent No. 5530189  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Ausich, Rodney L  
 ? APPLICANT: Brinkhaus, Friedrich L  
 ? APPLICANT: Mukharji, Indrani  
 ? APPLICANT: Proffitt, John H  
 ? APPLICANT: Vargier, James G  
 ? APPLICANT: Yen, Hwei-Chue B  
 ? TITLE OF INVENTION: Lycopene Biosynthesis in  
 ? TITLE OF INVENTION: Genetically Engineered Hosts  
 ? NUMBER OF SEQUENCES: 70  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESS: Amoco Corp., Patents and Licensing Dept  
 ? STREET: 200 E Randolph St  
 ?

CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60680-0703  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.24  
CURRENT APPLICATION NUMBER: US/08/096,043  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/785,568  
FILING DATE: 30-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, No. 5530189val B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 3128567180  
TELEFAX: 3128564972  
INFORMATION FOR SEQ. ID. NO. 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 308 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-096-043-6

Query Match

Query March	65.0%;	Score 997.5;	DB 1;	Length 308;
Best Local Similarity	83.9%;	Pred. N.6.3e-102;		
Matches 189;	Conservative 36;	Mismatches 70;	Indels 1;	Gaps 1;
QY	1	MAVGSKSFAFATSLFDPAKTRRSVLMIYAWCRCCDDVIDDQTILGFHADQFSQMPBQLQQ	60	
DB	14	MANGSKSFATAAKLFPDPATRRSVLMLYTWCRCCDDVIDDQTHGFASAAAEEEAQTQLAR	73	
QY	61	LEMKTRQAVAGSQHSPAPAAQEQVAMAHDIAPAYAFPHLEGFAMDRVETRYLTLLDDTLR	120	
DB	74	LRITLTAAFGAEKQDPAPAAQEQVALTHGITPRMALDHLDDGFAMDVAQTVVFEDTLR	133	
QY	121	YCYHVGAVGLMKAQITMGVRYDNATLDRACDLGLAQLTNIARDIVDDAQOVCRCVLPESWL	180	
DB	134	YCYHVGAVGLMKAARVMGVRYDERVLDRACDLGLAQLTNMARDIIDDAARDICVLPBAWL	193	
QY	181	EEERGLTKANYAAPENRQALSRJAGRLVREAEPPYVYSSNAGLAQLPLRSAMAIATAKQVYR	240	

b 194 QDAGLAPENYAARENRPALARWR-BLIDAAEPPYISSOAGLHDLRRRSAMAIATARSYR 252  
Y 241 KIGVKVEQAGKAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP1 296  
b 253 EIGIKVKAAGSANDRRQHTSGEKIAMLMAAPGQVIRAKTRVTPRPAGLWQRPV 308

RESULT 6  
3-08-093-577-6  
Sequence 6, Application US/08093577  
Patent No. 5545816  
GENERAL INFORMATION:  
APPLICANT: Ausich, Rodney L.  
APPLICANT: Brinkhaus, Friedhelm L.  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H.  
APPLICANT: Yarger, James G.  
APPLICANT: Yen, Huel-Che B.  
TITLE OF INVENTION: Phytoene Biosynthesis in  
TITLE OF INVENTION: Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
STREET: 200 E Randolph St  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60680-0703  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/093,577  
FILING DATE: 19-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/785,569  
FILING DATE: 30-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, No. 5545816val B  
TELEPHONE: 3128567180  
TELEFAX: 3128564972  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 308 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

S-08-093-577-6  
Query Match 65.0%; Score 997.5; DB 1; Length 308;  
Best Local Similarity 63.9%; Pred. No. 6.3e-102;  
Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;  
Y 1 MAVGSKSFATATLFDKTRRSVLMYAWCRHCDVDIDQTLGFHADQPSQMPERLQQ 60  
b 14 MANGSKSFATAKLFDKTRRSVLMYAWCRHCDVDIDQTLGFHADQPSQMPERLQQ 73  
Y 61 LEWTKQVAGSOMEPAPAPQFVAMAHDIAPAYAFHLEGFANDVETRYLTDTLR 120  
b 74 LRTLTAEEGAEMQDPAPAPQFVAMAHDIAPAYAFHLEGFANDVETRYLTDTLR 133  
Y 121 YCVHVAGVGLAWQIMGVDRNATIDRACDLGLAFQLTNIARDIVDAQVRCYLPESWL 180  
b 134 YCHVAGVGLAWQIMGVDRNATIDRACDLGLAFQLTNIARDIVDAQVRCYLPESWL 193  
Y 181 EREGTKYANAPENRQALSRAGRLVREAEPPYVSSMAGLAQLPLGSAMAIATAKOYR 240  
b 194 QDAGLAPENYAARENRPALARWR-BLIDAAEPPYISSOAGLHDLRRRSAMAIATARSYR 252

QY 241 KIGVKVEQAGKAWDHROSTSTAEKLTLLLTASGQAVTSRMKTYPPRPAHLWQRP1 296  
Db 253 EIGIKVKAAGSANDRRQHTSGEKIAMLMAAPGQVIRAKTRVTPRPAGLWQRPV 308

RESULT 7  
US-08-096-623A-6  
Sequence 6, Application US/08096623A  
Patent No. 5684238  
GENERAL INFORMATION:  
APPLICANT: Ausich, Rodney L.  
APPLICANT: Brinkhaus, Friedhelm L.  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H.  
APPLICANT: Yarger, James G.  
APPLICANT: Yen, Huel-Che B.  
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and  
TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Welsh & Katz, Ltd.  
STREET: 120 S. Riverside Plaza, 22nd Floor  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,623A  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,061  
FILING DATE: 09-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/662,921  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/562,674  
FILING DATE: 03-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/525,551  
FILING DATE: 18-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,613  
FILING DATE: 02-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Ganson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER: AMO-006.1  
TELEPHONE: (312) 655-1500  
TELEFAX: (312) 655-1501  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 308 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-096-623A-6

Query Match 65.0%; Score 997.5; DB 1; Length 308;  
Best Local Similarity 63.9%; Pred. No. 6.3e-102;  
Matches 189; Conservative 36; Mismatches 70; Indels 1; Gaps 1;  
QY 1 MAVGSKSFATATLFDKTRRSVLMYAWCRHCDVDIDQTLGFHADQPSQMPERLQQ 60  
Db 14 MANGSKSFATAKLFDKTRRSVLMYAWCRHCDVDIDQTLGFHADQPSQMPERLQQ 73





116 DDTLYCYHVAGVGLMAQINGVRDNATLDACDLGLAFQLTNIARDVDDAQVGRCYL 175  
125 DDVLEYSYHVAGVGVWMAVGVQDDAVLDRACDLGLAFQLTNIARDVDDAAIGRCYL 184  
176 PESWLEEGTLTKANYAAPENRQALSRIAGRLVREAPFYVSSMAGLAQLPLRSAAIATA 235  
185 PADWLAEG--ATVEGVPVSDALYSVIIRLLDAAEPYASARQGLPHLPFPCAWSIATA 241  
236 KOVYRKIGVKEQAGKQAWDHROSTSTAEBKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 295  
242 LRIYRAIGTRIRQGGPEAYRQISTSKAAKIGLLARGGLDAAASRLRGGEISRDGLWTRP 301

## RESULT 10

US-09-546-969-3

Sequence 3, Application US/09546969

Patent No. 6207409

## GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/09/546,969

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/660,645

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 303 amino acids

TYPES: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Protein

US-09-546-969-3

Query Match 44.7%; Score 686.5; DB 3; Length 303;

Best Local Similarity 48.7%; Pred. No. 1.6e-67;

Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;

1 MAVGSKSFATASTLFDKATRSVLMYAWCRHCDVDVDDQTLGFADQSSQMPQRLOQ 60  
12 IAQGSQSFQAQAKLMPGIREDTVLMYAWCRHADDVDDQVMSAPEAGD--PQARLGA 69  
61 LEMKTRQAYAGSQMHE-----PAPAFOEVAMADHAPAVAFHLEGFAMDVRETRYLTL 115  
70 LRADTLAA-----LHEDGPMSPFPALRQVARRHDFDLWPMDLIEGFAMDVADRYRSL 124  
116 DDTLYCYHVAGVGLMAQINGVRDNATLDACDLGLAFQLTNIARDVDDAQVGRCYL 175

125 DDVLEYSYHVAGVGVWMAVGVQDDAVLDRACDLGLAFQLTNIARDVDDAAIGRCYL 184  
176 PESWLEEGTLTKANYAAPENRQALSRIAGRLVREAPFYVSSMAGLAQLPLRSAAIATA 235  
185 PADWLAEG--ATVEGVPVSDALYSVIIRLLDAAEPYASARQGLPHLPFPCAWSIATA 241  
236 KOVYRKIGVKEQAGKQAWDHROSTSTAEBKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 295  
242 LRIYRAIGTRIRQGGPEAYRQISTSKAAKIGLLARGGLDAAASRLRGGEISRDGLWTRP 301

## RESULT 11

US-08-980-832-3

Sequence 3, Application US/08980832B

Patent No. 6291204

## GENERAL INFORMATION:

APPLICANT: Pasamontes, Luis

APPLICANT: Tsygankov, Yuri

TITLE OF INVENTION: Improved Fermentative Carotenoid Production

FILE REFERENCE: Improved Fermentative Carotenoid

CURRENT APPLICATION NUMBER: US/08/980,832B

CURRENT FILING DATE: 1997-12-01

NUMBER OF SEQ ID NOS: 66

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 3

LENGTH: 303

TYPE: PRT

ORGANISM: Flavobacterium sp. R1534

US-08-980-832-3

Query Match

44.7%; Score 686.5; DB 3; Length 303;

Best Local Similarity 48.7%; Pred. No. 1.6e-67;

Matches 146; Conservative 35; Mismatches 104; Indels 15; Gaps 4;

1 MAVGSKSFATASTLFDKATRSVLMYAWCRHCDVDVDDQTLGFADQSSQMPQRLOQ 60  
12 IAQGSQSFQAQAKLMPGIREDTVLMYAWCRHADDVDDQVMSAPEAGD--PQARLGA 69

61 LEMKTRQAYAGSQMHE-----PAPAFOEVAMADHAPAVAFHLEGFAMDVRETRYLTL 115  
70 LRADTLAA-----LHEDGPMSPFPALRQVARRHDFDLWPMDLIEGFAMDVADRYRSL 124

116 DDTLYCYHVAGVGLMAQINGVRDNATLDACDLGLAFQLTNIARDVDDAQVGRCYL 175  
125 DDVLEYSYHVAGVGVWMAVGVQDDAVLDRACDLGLAFQLTNIARDVDDAAIGRCYL 184

176 PESWLEEGTLTKANYAAPENRQALSRIAGRLVREAPFYVSSMAGLAQLPLRSAAIATA 235  
185 PADWLAEG--ATVEGVPVSDALYSVIIRLLDAAEPYASARQGLPHLPFPCAWSIATA 241

236 KOVYRKIGVKEQAGKQAWDHROSTSTAEBKLTLLLTASQAVTSRMKTYPPRPAHLWQRP 295  
242 LRIYRAIGTRIRQGGPEAYRQISTSKAAKIGLLARGGLDAAASRLRGGEISRDGLWTRP 301

## RESULT 12

US-09-547-267-3

Sequence 3, Application US/09547267

Patent No. 6613543

## GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

62 ALEALFTLAGRMEDYVRLKFLVFKFYLWCWESGKTLTDYANDSFELEAQITQFNPRDVEGY 121  
127 QRFIDYSKNLCTEAGFAAG---LDGFMDLLKFGYGLPLSLLSFDVFRSMDQVRRFIS 183  
122 RFLAYSAVQFQ---GYRLGSPVFLSFRDMLR-AGP---QLLKQAWQSVYQSVSRFIE 175  
184 DPKLVEILNYFIKVVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELG 243  
176 DEHLRQAFSPHLLVGNPFTTSIYTLIHALERENGWVPEGCTGALVNGMWKLTDLG 235  
244 VEIRLDAEVSIOQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLARS-PASELK-- 300  
236 GEIELNARVEELVADVNRVSVRLADGRIFDTDAVSNADVNTYKLLGTIPVQGRRA 295  
301 KQRFPEPSCGLVHLGVDRILYPOLAHNPFYSDPHREHDAVFKSHRLSDDDPTIYLVA 360  
296 RLERKSMNSLFLVLYFGLNQPHSLAHTTICFGPRYRELIDEIFTGSALADDPFSLYHSP 355  
361 CKTDPAPACACEIKILPHIPHLDPDKLLTAEDYSALRERVLVKKERGLTDLRQHV 420  
356 CVTDPSTLAPPFCASFYVLAPVPHLGNAPLDAQGPGLRDRIFDYLEERYMPLRSOLVT 415  
421 BEYMTPLDIQAKYYSNOGSY-----GVVADRFRKNLGFAPQRSSELSNLYFVGGSV 472  
416 QRIFTQTSRHWAIILGSLFIEPPSLTQGLFA-----ANATRSNLYLVAAGT 464  
473 NPGGMPMTLSGQVLRDKIVADLQ 497  
465 HPGAGIPGWGLAESTASLTIEDLQ 489

## RESULT 7

-08-096-623A-8  
Sequence 8, Application US/08096623A  
Patent No. 5684238  
GENERAL INFORMATION:  
APPLICANT: Ausich, Rodney L.  
APPLICANT: Brinkhaus, Friedhelm L.  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H.  
APPLICANT: Yarger, James G.  
APPLICANT: Yen, Hwei-Chue B.  
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and  
TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Weish & Katz, Ltd.  
STREET: 120 S. Riverside Plaza, 22nd Floor  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096.623A  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,061  
FILING DATE: 09-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/662,921  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/562,674  
FILING DATE: 03-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/525,551

FILING DATE: 18-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,613  
FILING DATE: 02-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Gamson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER: AMO-006.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 655-1500  
TELEFAX: (312) 655-1501  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 489 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-096-623A-8

Query Match 27.0%; Score 702.5; DB 1; Length 489;

Best local Similarity 32.9%; Pred. No. 1.9e-63;

Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

QY 7 QRVIVIGAGLGLSAAISLATAGFSVLIRKNDKVGKLNIMTKDGTFFDLGSPSILTMPH 66  
DB 2 KKTVVIGAGFGGLAALRLQAAGIPTVLLEQRDKPGGRAYVWHQDQGTFFDAGPTVITDPT 61  
QY 67 IFEALFTGAGNMADYVOIQKVEPHNRNFFEDGSDVIDCEDAETQRRELDKLGCTVAQF 126  
DB 62 ALEALFTLAGRMEDYVRLKFLVFKFYLWCWESGKTLTDYANDSFELEAQITQFNPRDVEGY 121  
QY 127 QRFIDYSKNLCTEAGFAAG---LDGFMDLLKFGYGLPLSLLSFDVFRSMDQVRRFIS 183  
DB 122 RFLAYSAVQFQ---GYRLGSPVFLSFRDMLR-AGP---QLLKQAWQSVYQSVSRFIE 175  
QY 184 DPKLVEILNYFIKVVGSPPYDAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELG 243  
DB 176 DEHLRQAFSPHLLVGNPFTTSIYTLIHALERENGWVPEGCTGALVNGMWKLTDLG 235  
QY 244 VEIRLDAEVSIOQDGRACAVKLANGDVLPAIVVSNMEVIPAMEKLARS-PASELK-- 300  
DB 236 GEIELNARVEELVADVNRVSVRLADGRIFDTDAVSNADVNTYKLLGTIPVQGRRA 295  
QY 301 KQRFPEPSCGLVHLGVDRILYPOLAHNPFYSDPHREHDAVFKSHRLSDDDPTIYLVA 360  
DB 296 RLERKSMNSLFLVLYFGLNQPHSLAHTTICFGPRYRELIDEIFTGSALADDPFSLYHSP 355  
QY 361 CKTDPAPACACEIKILPHIPHLDPDKLLTAEDYSALRERVLVKKERGLTDLRQHV 420  
DB 356 CVTDPSTLAPPFCASFYVLAPVPHLGNAPLDAQGPGLRDRIFDYLEERYMPLRSOLVT 415  
QY 421 BEYMTPLDIQAKYYSNOGSY-----GVVADRFRKNLGFAPQRSSELSNLYFVGGSV 472  
DB 416 QRIFTQTSRHWAIILGSLFIEPPSLTQGLFA-----ANATRSNLYLVAAGT 464  
QY 473 NPGGMPMTLSGQVLRDKIVADLQ 497  
DB 465 HPGAGIPGWGLAESTASLTIEDLQ 489

## RESULT 8

US-07-783-705A-4  
Sequence 4, Application US/07783705A  
Patent No. 5429939  
GENERAL INFORMATION:  
APPLICANT: Misawa, No. 5429939ihiko  
APPLICANT: Kobayashi, Kazuo  
APPLICANT: Nakamura, Katsumi  
APPLICANT: Yamano, Shigeyuki  
TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE  
TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS  
NUMBER OF SEQUENCES: 18

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Ladas & Parry  
STREET: 26 West 61 Street  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10023

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: N/A

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/783,705A  
FILING DATE: 19911023  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 1-103078  
FILING DATE: 21-APR-1989  
APPLICATION NUMBER: JP 2-53225  
FILING DATE: 05-MAR-1990  
APPLICATION NUMBER: US 07/519,011  
FILING DATE: 19-APR-1990

## ATTORNEY/AGENT INFORMATION:

NAME: Schwadron, Janet I.  
REGISTRATION NUMBER: 33,778  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-708-1935  
TELEFAX: 212-246-5959

## INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 492 amino acids  
TYPE: AMINO ACID

## TOPOLOGY: linear

## MOLECULE TYPE: protein

IS-07-783-705A-4

Query Match 25.8%; Score 671; DB 1; Length 492;

Best Local Similarity 32.7%; Pred. No. 3.3e-60;

Matches 161; Conservative 89; Mismatches 230; Indels 12; Gaps 5;

```

2Y 11 VIGAGLGLSALISLATAGFSVQLIEKNKDKVGGKLNMTKDGFTDLGSPILTMPIHFE 70
2b ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2Y 6 VIGAGFGLALAIRLOAGIPVLLERQDKGGRAYVYEDQGFTHDGPVITVDPISAIEE 65
2b ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2Y 71 LFTGAKNMADYVQIQKVEPHWRNFFEDGSDVIDLCEDAETQRRELKLGPGTYAQORFL 130
2b ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2b 66 LPALAGKQLXEVRELLPVPTEYRLCWBSGKVFYNDNDQTRLEAQIQQFNPRDVEGYRQL 125
2Y 131 DYSKNLCETETAGYPAKGLDGFMDLLKFGYGLRS---LLSPDVFRSMDQVRSFISDPKL 187
2b ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2b 126 DISRAVFE---GYLKLGTVPF---LSFRMLRAAPQLAKLQARSVYSKVASYIEDEHL 179
2Y 188 VEILNYFIKYVGSSPYDAPALMNLPIYQHYGLWYKGMYGMAQAMEKLAVELGVEIR 247
2b ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2b 180 ROAFSGHLLVGNPFPATSSITLIHALEREWGVPFRGGTGALVQGMKLPDLGGEVV 239
2Y 248 LDAEYSEIQKQDGRACAVKLANGDVLPAIVVSNMEVIPAAMEKLL-RSPAS--ELKGMQR 304
2b ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2b 240 LNAVSEMTTGNKIEAVLEEDGRFTQAVASNADVVHTYRDLSSQHPARKVQSNLGT 299
2Y 305 FPPSCGLVTLGLVDRLYPLAHENFYSDHPRSHFDVAFKSHRLSDPTIYLVAPCKTD 364
2b ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2b 300 KMSNSLFLVYFGLMHHDDQLAHTVCGFGRPRYRELIDEIFNHDLAEDFSLYLHAPCVTD 359
2Y 365 PAQAPAGCEIILKILPHLDPDKLLTAEDYSALRERVLKLERMGLTDLRCHIVTEEYV 424
2b ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2b 360 SSLAPEGCGSYVLAAPVHLGTANLDTVEGPKLRDIFAYLEQHTNPGRLQVTHRMF 419
2Y 425 TELDIQAKYNSQGSYGVWADRFKGLGFAKQPSSELSNLYFVGVSVPFGGMPVTLIS 484
2b ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2b 420 TPFDPDQLNHYGSAFSEVPVLQSAFNPFRNDRKTIITNLVVGAGTHPGAGIGVIGS 479

```

QY 485 GQLVRDKIVADL 496  
Db 480 AKATAGLMLDEL 491

## RESULT 9

US-08-660-645A-5

Sequence 5, Application US/08660645A

Patent No. 6087152

GENERAL INFORMATION:

APPLICANT: Hohmann, Hans-Peter

APPLICANT: Pasamontes, Luis

APPLICANT: Tessier, Michel

APPLICANT: van Loon, Adolphus

TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hoffmann-La Roche Inc.

STREET: 340 Kingsland Street

CITY: Nutley

STATE: NJ

COUNTRY: USA

ZIP: 07110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/660,645A

FILING DATE: 07-JUN-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 9510888.9

FILING DATE: 09-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Pokras, Bruce A.

REGISTRATION NUMBER: 32,748

REFERENCE/DOCKET NUMBER: RAN 6002/170

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-5801

TELEFAX: (201) 235-2363

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 494 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-660-645A-5

Query Match 24.5%; Score 637; DB 3; Length 494;

Best Local Similarity 31.2%; Pred. No. 1.1e-56;

Matches 155; Conservative 90; Mismatches 233; Indels 18; Gaps 6;

```

QY 10 IVIGAGLGLSALISLATAGFSVQLIEKNKDKVGGKLNMTKDGFTDLGSPILTMPIHFE 69
Db 5 IVIGAGFGLALAIRLOAGIATTVIARDKFGGRAYVNDQGHVFDAGPTVTDPSLR 64
QY 70 ALPTGAKNMADYVQIQKVEPHWRNFFEDGSDVIDLCEDAETQRRELKLGPGTYAQORF 129
Db 65 ELWALSGQPMERDVTLLPVSPFFRLTWADGRSFETVNDDELIRQVASFNFADVDGYRRF 124
QY 130 LDYSKNLCETETAGYPAKGLDGFMDLLKFGYGLRSLLSFDVFRSMDQVRRFISDPKLVE 189
Db 125 HDYAEVYRE---GYLKLGTTPFLKGLQMLNAAAPALMLQAVRSVHSMVAFIQQDPLHQ 181
QY 190 ILNYPFKYVGSSPYDAPALMNLPIYQHYGLWYKGMYGMAQAMEKLAVELGVEIRLD 249
Db 182 AFSFHTLLVGNPFPSTSSIIYALIHALERGGVWFRKGTGNOLVAGMVALFERLGGTLLLN 241
QY 250 AEYSEIQKQDGRACAVKLANGDVLPAIVVSNMEVIPAAMEKLL-----RSPASELKMQ 303
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

b 242 ARVTRIDTEGDRATGVTLLDGRQLRADTVASNGDVHSHYRDLGHTRRGRTKAAILNR-Q 300  
y 304 RFEPCSGVLHLGVDRLYPQLAHNFFYSDHPRHFDVAFKSHLSDDPTIYLVAPOKT 363  
b 301 RW--SMSFLVHLFGSKRPENLAHSHVIGPRYKGLVNEIFNGPRLPDDESMYLHSPCVT 358  
y 364 DPAQAPAGCEIIKILPHIEL---DPDKLLTAEDYSALRERVVLKLERMGLTDLRQHIVT 420  
b 359 DPLAPEGSTHYVLAPVPHLGRADVWEAEPGYA---ERIFESLERRAIPDLRKHIVT 415  
y 421 EEWYTPDLIOAKYYSNOGSIYGVVADRFRKNGFKAPQRSSELSNLYFVGGSNPGGMPM 480  
b 416 SRIFSPADFSTELSAHSGAFSVEPILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPG 475  
y 481 VTLSGQLVRDKIVADL 496  
b 476 VVGSAKATAQVMLSD 491

## RESULT 10

S-09-298-718-5  
Sequence 5, Application US/09298718  
Patent No. 6124113  
GENERAL INFORMATION:  
APPLICANT: Hohmann, Hans-Peter  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tessier, Michel  
APPLICANT: van Loon, Adolphus  
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: NJ  
COUNTRY: USA  
ZIP: 07110

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/298,718  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/660,645  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pokras, Bruce A.  
REGISTRATION NUMBER: 32,748  
REFERENCE/DOCKET NUMBER: RAN 6002/170  
TELEPHONE: (201) 235-5801  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

## S-09-298-718-5

Query Match 24.5%; Score 637; DB 3; Length 494;  
Best Local Similarity 31.2%; Pred. No. 1.1e-56;  
Matches 155; Conservative 90; Mismatches 233; Indels 18; Gaps 6;  
y 10 IVIGAGLGLSALISATAGFSVQLIEKDKYGVKLNINIKDGFTDLGSPSLTMTPLHFE 69  
b 5 IVIGAGGGLALAIRLQSGAATTIVEARDKPGGRAYVWVDQGHVFDAGPTVTDPSLR 64

Qy 70 ALFTGAGKWDYVQIQYEPHNRFFEDSGVIDLCEDAETORRELDKLGPGTYAQFORF 129  
Db 65 ELWALSGQPMERDVTLLPSPFFYRLTWADGRSFYVNDDELIQVASFNPADVDGRRF 124  
Qy 130 LDYSKNLCTETEAQYFAKGLDGFWDLLKFGPLRSLSPDFRSMKQGVRRFISDPKLYE 189  
Db 125 HDYABEVYRE--GYLKLGTTPFLKLGOMLNAAPALRLQAYRSVHSMVARPIQDPHLRQ 181  
Qy 190 ILNYFKYVCGSSYPDAPALMNLFPYQHYGLVYKGMVYGMQAQMEKLAVELGVIRLD 249  
Db 182 AFGFHTLLVGGNPFSTSIYALIHALLEREGGVWFAGGNTQNLVAGMVALFERLGGTLLN 241  
Qy 250 ABYSEITQKQGRACAVKLANGDVLPAIDIVVSNWEVIPAMEKLL-----RSPASELKQOQ 303  
Db 242 ARVTRIDTEGDRATGVTLLDGRQLRADTVASNGDVHSHYRDLGHTRRGRTKAAILNR-Q 300  
Qy 304 RFEPCSGVLHLGVDRLYPQLAHNFFYSDHPRHFDVAFKSHLSDDPTIYLVAPOKT 363  
Db 301 RW--SMSFLVHLFGSKRPENLAHSHVIGPRYKGLVNEIFNGPRLPDDESMYLHSPCVT 358  
Qy 364 DPAQAPAGCEIIKILPHIEL---DPDKLLTAEDYSALRERVVLKLERMGLTDLRQHIVT 420  
Db 359 DPLAPEGSTHYVLAPVPHLGRADVWEAEPGYA---ERIFESLERRAIPDLRKHIVT 415  
Qy 421 EEWYTPDLIOAKYYSNOGSIYGVVADRFRKNGFKAPQRSSELSNLYFVGGSNPGGMPM 480  
Db 416 SRIFSPADFSTELSAHSGAFSVEPILTQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPG 475  
Qy 481 VTLSGQLVRDKIVADL 496  
Db 476 VVGSAKATAQVMLSD 491

## RESULT 11

US-09-546-969-5  
Sequence 5, Application US/09546969  
Patent No. 620409  
GENERAL INFORMATION:  
APPLICANT: Hohmann, Hans-Peter  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tessier, Michel  
APPLICANT: van Loon, Adolphus  
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: NJ  
COUNTRY: USA  
ZIP: 07110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/546,969  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/660,645  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pokras, Bruce A.  
REGISTRATION NUMBER: 32,748  
REFERENCE/DOCKET NUMBER: RAN 6002/170  
TELEPHONE: (201) 235-5801  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids



LENGTH: 494 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
-09-547-267-5

Query Match 24.5%; Score 637; DB 4; Length 494;

Best Local Similarity 31.2%; Pred. No. 1.1e-56;

Matches 155; Conservative 90; Mismatches 233; Indels 18; Gaps 6;

```
10 IVIGAGLGLSAAISLATAGSVOLIEKNKVKGKLNIMTKDGTFTDLGPSILTMHPHIF 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5 IVIGAGLGLSAAISLATAGSVOLIEKNKVKGKLNIMTKDGTFTDLGPSILTMHPHIF 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
70 ALFTGAGNMADYVQIOKVEPHNFFEDSGSVIDLCEDAETORRELDKLGPGTYAOFQRF 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65 ELWALSQGMERDVTLLPVSPFYRLTWADGRSFEYVNDDELIRQVASFNPADVDGYRRF 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 LDYSKNLCTETEGYFAKGLDGFWDLLKFTGPIRLSLLSPDVFRRSMQGVRRFISDPKIVE 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
125 HDYAEVYRE---GYLKLGTTPFLKLGQMLNAAPALMRLQAYRSVHSMVARFIQDPHLRQ 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
190 ILNFIKTVGSSPDAPALMNLPIYIOHYGLVWVGKSYGMAQAMEKLAVELGVEIRLD 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 AFSFHTLLVGNPFSTSSIVAIHALLERGGVWFAKGTNQLVAGVVALPERLGGTLLLN 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 AEVSEIQKQGRACAVKLANGDVLPAIVVSNMEVIPAMEKLL-----RSPASELKXMQ 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
242 ARVTIDTEGDRATGVTLLDGRQLRADTVASNGDVMSYRDLLGHTRRGRTKAAILNR-Q 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
304 RFEPSCSGLVHLGVDRILYPOLAHNFFYSDHPREHFDVAFKSHRLSDDPTIYLVAPCKT 363
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 RW--SMSLFVLHFLGSKRPENLAHSHVIFGPRYKGLVNEIFNGRPLPDDFSMILHSPCVT 358
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 DPAQAPACEIIKILPHIPLH---DPDKLLTAEDYSALRERVLKLERMGLTDLRQHIVT 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
359 DPLSAPEGMSTHYVLAPVPHLGRADVWEAEPGYA---ERIFELEERRAIPDLRKHGLTV 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
421 BEYWTPLDIQAKYNSQGSIVGVVADRFKNGKAPORSSELSNLYFVGSGVNPGGMPM 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
416 SRIFSPADFSTELSAHSGSAFSEPIITQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPG 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
481 VTLSQGLVRDKIVADL 496
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
476 VVGSAKATAQVWLSDL 491
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 14

-09-920-923B-4

Sequence 4, Application US/09920923B

Patent No. 6571134

GENERAL INFORMATION:

APPLICANT: Pasamontes, Luis

APPLICANT: Tsygankov, Yuri

TITLE OF INVENTION: Fermentative Carotenoid Production

FILE REFERENCE: 15464 US (C38435/125944)

CURRENT APPLICATION NUMBER: US/09/920,923B

CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 08/980,832

PRIOR FILING DATE: 1997-12-01

NUMBER OF SEQ ID NOS: 66

SOFTWARE: Patent in version 3.1

SEQ ID NO 4

LENGTH: 494

TYPE: PR

ORGANISM: Flavobacterium sp. R1534

-09-920-923B-4

Query Match

Best Local Similarity 24.5%; Score 637; DB 4; Length 494;

Matches 155; Conservative 90; Mismatches 233; Indels 18; Gaps 6;

```
QY 10 IVIGAGLGLSAAISLATAGSVOLIEKNKVKGKLNIMTKDGTFTDLGPSILTMHPHIF 69
Db 5 IVIGAGLGLSAAISLATAGSVOLIEKNKVKGKLNIMTKDGTFTDLGPSILTMHPHIF 64
QY 70 ALFTGAGNMADYVQIOKVEPHNFFEDSGSVIDLCEDAETORRELDKLGPGTYAOFQRF 129
Db 65 ELWALSQGMERDVTLLPVSPFYRLTWADGRSFEYVNDDELIRQVASFNPADVDGYRRF 124
QY 130 LDYSKNLCTETEGYFAKGLDGFWDLLKFTGPIRLSLLSPDVFRRSMQGVRRFISDPKIVE 189
Db 125 HDYAEVYRE---GYLKLGTTPFLKLGQMLNAAPALMRLQAYRSVHSMVARFIQDPHLRQ 181
QY 190 ILNFIKTVGSSPDAPALMNLPIYIOHYGLVWVGKSYGMAQAMEKLAVELGVEIRLD 249
Db 182 AFSFHTLLVGNPFSTSSIVAIHALLERGGVWFAKGTNQLVAGVVALPERLGGTLLLN 241
QY 250 AEVSEIQKQGRACAVKLANGDVLPAIVVSNMEVIPAMEKLL-----RSPASELKXMQ 303
Db 242 ARVTIDTEGDRATGVTLLDGRQLRADTVASNGDVMSYRDLLGHTRRGRTKAAILNR-Q 300
QY 304 RFEPSCSGLVHLGVDRILYPOLAHNFFYSDHPREHFDVAFKSHRLSDDPTIYLVAPCKT 363
Db 301 RW--SMSLFVLHFLGSKRPENLAHSHVIFGPRYKGLVNEIFNGRPLPDDFSMILHSPCVT 358
QY 364 DPAQAPACEIIKILPHIPLH---DPDKLLTAEDYSALRERVLKLERMGLTDLRQHIVT 420
Db 359 DPLSAPEGMSTHYVLAPVPHLGRADVWEAEPGYA---ERIFELEERRAIPDLRKHGLTV 415
QY 421 BEYWTPLDIQAKYNSQGSIVGVVADRFKNGKAPORSSELSNLYFVGSGVNPGGMPM 480
Db 416 SRIFSPADFSTELSAHSGSAFSEPIITQSAWFRPHNRDRAIPNFYIVGAGTHPGAGIPG 475
QY 481 VTLSQGLVRDKIVADL 496
Db 476 VVGSAKATAQVWLSDL 491
```

RESULT 15

US-09-091-725-17

Sequence 17, Application US/09091725

Patent No. 6329141

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Improved methods for transforming Phaffia

TITLE OF INVENTION: and recombinant DNA for use therein

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morrison & Foerster LLP

STREET: 2000 Pennsylvania Avenue, N.W.

CITY: Washington

STATE: DC

COUNTRY: United States of America

ZIP: 20006-1898

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25 (BPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/091,725

FILING DATE: 23-DEC-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 95203620.0

FILING DATE: 22-DEC-1995

APPLICATION NUMBER: EP 96200943.7

FILING DATE: 11-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: E. Victor Donahue

REGISTRATION NUMBER: 35,492

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 582 amino acids

Search completed: February 29, 2004, 14:54:53  
Job time : 17.1029 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:36:38 ; Search time 58.7617 Seconds  
(without alignments)  
2389.754 Million cell updates/sec

file: US-09-941-947A-24

effect score: 2598  
sequence: 1 MNSNDNQRVIVIGAGLGGLS.....MPMTLSQQLVRDKIVADLQ 497

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

sarched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

imum DB seq length: 0  
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- atabase :
- 1: A\_Geneseq\_29Jan04:\*
  - 2: Geneseq1980s:\*
  - 3: Geneseq1990s:\*
  - 4: Geneseq2000s:\*
  - 5: Geneseq2001s:\*
  - 6: Geneseq2002s:\*
  - 7: Geneseq2003as:\*
  - 8: Geneseq2003bs:\*
  - 9: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	ID	Description
1	2598	100.0	497	5	Aae22310 Methylomo
2	2598	100.0	497	5	Abg61589 High grow
3	2598	100.0	497	5	Aau80333 Methylomo
4	2598	100.0	497	6	Ada14538 Methylomo
5	1289	49.6	497	6	Ada14544 Staphyloc
6	1289	49.6	500	6	Abm72616 Staphyloc
7	1287	49.5	497	6	Abu16031 Protein e
8	1280	49.3	502	4	Aau34200 Staphyloc
9	1039	40.0	439	4	Aau34200 Staphyloc
10	713	27.4	494	6	Abm70123 Photorhab
11	703.5	27.1	489	2	Aar13985 Phytoene
12	703.5	27.1	489	2	Aar13984 Phytoene
13	703.5	27.1	489	2	Aaw01123 Phytoene
14	703.5	27.1	489	2	Aaw00174 Phytoene
15	703.5	27.1	489	2	Aaw32475 Erwinia h
16	702.5	27.0	489	2	Aaw01122 Phytoene
17	702.5	27.0	489	2	Aaw00173 Phytoene
18	702.5	27.0	489	2	Aaw32472 Erwinia h
19	687	26.4	492	5	Aae22314 Pantoea s
20	687	26.4	492	6	Aao16021 Pantoea s
21	687	26.4	492	6	Abp96688 Pantoea s
22	672	25.9	502	6	Abm72613 Staphyloc
23	672	25.9	502	6	Ada14542 Staphyloc
24	671	25.8	492	2	Aar07466 Polypepti
25	671	25.8	492	2	Aaw82257 C. utilis

26	671	25.8	492	2	Aaw87889 Protein e
27	671	25.8	492	2	Aaw99099 Erwinia u
28	671	25.8	492	2	Aay26333 Erwinia u
29	653	25.1	499	2	Aar64269 Anabaena
30	640.5	24.7	582	6	Abp97465 Blakeslea
31	637	24.5	494	2	Aaw00871 Flavobact
32	634	24.4	494	2	Aaw69532 Flavobact
33	618.5	23.8	582	2	Aaw22499 Phaffia d
34	607.5	23.4	511	5	Aae22309 Methylomo
35	607.5	23.4	511	5	Abg61588 High grow
36	607.5	23.4	511	5	Aau80332 Methylomo
37	607.5	23.4	511	6	Ada14534 Methylomo
38	596	22.9	290	2	Aaw77702 Crtd proc
39	558	21.5	526	2	Aar95697 Erythroba
40	554	21.3	490	6	Abu32531 Protein e
41	551	21.2	490	5	Abb49224 Listeria
42	522	20.1	543	5	Aao15521 Micrococc
43	495	19.1	544	5	Aao15518 Agromyces
44	475	18.3	548	4	Aab76640 Corynebact
45	475	18.3	548	4	Aab76641 Corynebact

ALIGNMENTS

RESULT 1  
AAE22310 standard; protein: 497 AA.

XX AAE22310;  
AC  
XX  
XX  
DT 07-JUL-2003 (revised)  
DT 25-JUL-2002 (first entry)  
XX  
XX Methylomonas 16a sp. diapophytoene dehydrogenase (Crtn2) enzyme.  
XX Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet;  
KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;  
KW aquaculture; enzyme; diapophytoene dehydrogenase; Crtn2.  
XX Methylomonas sp.  
XX WO200218617-A2.  
XX  
XX  
PD 07-MAR-2002.  
XX  
PF 04-SEP-2001; 2001WO-05027490.  
XX  
PR 01-SEP-2000; 2000US-0229858P.  
PR 01-SEP-2000; 2000US-0229907P.  
XX  
XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX  
XX Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;  
XX Odom JM, Picataggio SK, Rouviere PE;  
XX  
XX WPI; 2002-351711/38.  
XX N-PSDB; AAD35508.  
XX  
XX Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by  
XX using microorganisms having a nucleic acid molecule encoding enzymes in  
XX the carotenoid biosynthetic pathway and which metabolize single carbon  
XX substrates.  
XX  
XX Claim 37; Page 131-133; 156pp; English.  
XX  
XX The invention relates to a method for producing carotenoid compounds. The  
XX method comprises a transformed metabolising host cell, comprising  
XX suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule  
XX encoding an enzyme in the carotenoid biosynthetic pathway, under the  
XX control of regulatory sequences, and contacting the host cell with carbon  
XX substrate to produce a carotenoid compound. The method is useful for  
XX producing carotenoid compounds such as anthraxanthin and astaxanthin, by



using microorganism having a nucleic acid molecule encoding enzymes in the carotenoid biosynthetic pathway and which metabolise single carbon substrates. The carotenoids have potent anti-oxidant properties useful in diet, and aquaculture elements. The carotenoids are also useful as intermediates in the synthesis of steroids flavours and fragrances and compounds for potential electro-optic applications. The present sequence is *Methylobacterium* 16a sp. diaphorocene dehydrogenase (CrtN2) enzyme used in the invention. (Updated on 07-AUG-2003 to correct OS field.)

XX	Sequence 497 AA;
XX	Query Match 100.0%; Score 2598; DB 5; Length 497;
XX	Best Local Similarity 100.0%; Pred. No. 1.5e-241;
XX	Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	1 MNSNDNRVIVIGAGLGGLSAAISLATAGFSVOLIERKNDKVGKLMIMTKDGTFFDLGSP 60
XX	1 MNSNDNRVIVIGAGLGGLSAAISLATAGFSVOLIERKNDKVGKLMIMTKDGTFFDLGSP 60
XX	61 ILTMPHIFEALFTGAGKNNMADYVQIQKVEPHNRNFFEDGSDVIDLCEAETORRELDKLP 120
XX	61 ILTMPHIFEALFTGAGKNNMADYVQIQKVEPHNRNFFEDGSDVIDLCEAETORRELDKLP 120
XX	121 GTYAQFORFLDYSKNLCTETETAGYFAKGLDGFWDLLKFYGLPLSLLSFDVFRSMQGVRR 180
XX	121 GTYAQFORFLDYSKNLCTETETAGYFAKGLDGFWDLLKFYGLPLSLLSFDVFRSMQGVRR 180
XX	181 FISDPKLVIELNYFIKVGSSPYDAPALMNLPIQVHYGLVYKGMGCAAMEKLAIV 240
XX	181 FISDPKLVIELNYFIKVGSSPYDAPALMNLPIQVHYGLVYKGMGCAAMEKLAIV 240
XX	241 ELGVEIRLDAEVSIIQKQDGRACAVKLANGDVLPAIVWSNMEVIPAAMEKLLSPASELK 300
XX	241 ELGVEIRLDAEVSIIQKQDGRACAVKLANGDVLPAIVWSNMEVIPAAMEKLLSPASELK 300
XX	301 KMORFEPSCGLVHLGVDRILYPOLAHNFFYSDHPREHFDVFKSHRLSDDDTIYLVAP 360
XX	301 KMORFEPSCGLVHLGVDRILYPOLAHNFFYSDHPREHFDVFKSHRLSDDDTIYLVAP 360
XX	361 CKTDPAPACACIIKILPHIPHLDPDKLLTABDYSALRERVLVKLERMGLTDLRQHI 420
XX	361 CKTDPAPACACIIKILPHIPHLDPDKLLTABDYSALRERVLVKLERMGLTDLRQHI 420
XX	421 EYWTPLDIQKYSNCGSIYGVVADRFKNGLFKAPQRSSELSNLYFVGSSVNPGGMPM 480
XX	421 EYWTPLDIQKYSNCGSIYGVVADRFKNGLFKAPQRSSELSNLYFVGSSVNPGGMPM 480
XX	481 VTLSQLVRDKIVADLQ 497
XX	481 VTLSQLVRDKIVADLQ 497

RESULT 2  
ABG61589  
ID ABG61589 standard; protein; 497 AA.  
XX ABG61589;  
XX ABG61589;  
DT 07-AUG-2003 (revised)  
DT 27-AUG-2002 (first entry)  
XX High growth methanotrophic bacterial strain polypeptide #39.  
XX High growth methanotrophic bacterial strain; C1 carbon substrate; enzyme;  
KW methane; methanol; Embden-Meyerhof carbon flux pathway; 16S RNA;  
KW pyrophosphate dependent phosphofructokinase; nitrogen-containing compound;  
KW ammonia; nitrate; nitrite; nitrogen; pigment; oxygen; landfill;  
KW methane-containing environment; waste water treatment system; isoprenoid;  
KW nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide.  
XX *Methylobacterium* sp.  
XX WO200220728-A2.  
PN

XX	14-MAR-2002.
XX	28-AUG-2001; 2001WO-US026827.
XX	01-SEP-2000; 2000US-0229858P.
XX	(DUPO) DU PONT DE NEMOURS & CO E I.
XX	Koffas M, Odom JM, Schenzle A;
XX	WPI; 2002-454358/48.
XX	N-PSDB; ABK83268.
XX	New high growth methanotrophic bacterial strain, useful for producing single cell proteins, grows on a C1 carbon substrate, and comprises a functional gene encoding in Embden-Meyerhof carbon pathway.
XX	Claim 11; Page 153-154; 157pp; English.
XX	The invention relates to a high growth methanotrophic bacterial strain, which grows on a C1 carbon substrate e.g. methane and methanol, and comprises a functional Embden-Meyerhof carbon flux pathway comprising a gene coding a pyrophosphate dependent phosphofructokinase enzyme or a 16S RNA. The bacterial strain is useful for the production of single cell protein and for the biotransformation of a nitrogen-containing compound, e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the production of a feed product comprising a protein, carbohydrates and a pigment and for reducing oxygen demand, for removing nitrates and nitrites in methane-containing environments such as landfills, waste water treatment systems or anywhere the invention can be used as a present. The bacterial strain of the invention can be used as a denitrifying agent for the conversion of nitrate or nitrite to nitrous oxide with methane or methanol as a carbon source. It is also used in the production of biomass including proteins, carbohydrates and a wide variety of pigments (particularly for isoprenoid pigments for the purpose of generating animal feeds), in production of terpenoid and carotenoid compounds, useful as pigments and as monomers in polymeric materials and in production of exopolysaccharides at high levels. Sequences ABG61551-ABG61590 represent high growth methanotrophic bacterial strain proteins of the invention. (Updated on 07-AUG-2003 to correct OS field.)
XX	Sequence 497 AA;
XX	Query Match 100.0%; Score 2598; DB 5; Length 497;
XX	Best Local Similarity 100.0%; Pred. No. 1.5e-241;
XX	Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	1 MNSNDNRVIVIGAGLGGLSAAISLATAGFSVOLIERKNDKVGKLMIMTKDGTFFDLGSP 60
XX	1 MNSNDNRVIVIGAGLGGLSAAISLATAGFSVOLIERKNDKVGKLMIMTKDGTFFDLGSP 60
XX	61 ILTMPHIFEALFTGAGKNNMADYVQIQKVEPHNRNFFEDGSDVIDLCEAETORRELDKLP 120
XX	61 ILTMPHIFEALFTGAGKNNMADYVQIQKVEPHNRNFFEDGSDVIDLCEAETORRELDKLP 120
XX	121 GTYAQFORFLDYSKNLCTETETAGYFAKGLDGFWDLLKFYGLPLSLLSFDVFRSMQGVRR 180
XX	121 GTYAQFORFLDYSKNLCTETETAGYFAKGLDGFWDLLKFYGLPLSLLSFDVFRSMQGVRR 180
XX	181 FISDPKLVIELNYFIKVGSSPYDAPALMNLPIQVHYGLVYKGMGCAAMEKLAIV 240
XX	181 FISDPKLVIELNYFIKVGSSPYDAPALMNLPIQVHYGLVYKGMGCAAMEKLAIV 240
XX	241 ELGVEIRLDAEVSIIQKQDGRACAVKLANGDVLPAIVWSNMEVIPAAMEKLLSPASELK 300
XX	241 ELGVEIRLDAEVSIIQKQDGRACAVKLANGDVLPAIVWSNMEVIPAAMEKLLSPASELK 300
XX	301 KMORFEPSCGLVHLGVDRILYPOLAHNFFYSDHPREHFDVFKSHRLSDDDTIYLVAP 360
XX	301 KMORFEPSCGLVHLGVDRILYPOLAHNFFYSDHPREHFDVFKSHRLSDDDTIYLVAP 360
XX	361 CKTDPAPACACIIKILPHIPHLDPDKLLTABDYSALRERVLVKLERMGLTDLRQHI 420
XX	361 CKTDPAPACACIIKILPHIPHLDPDKLLTABDYSALRERVLVKLERMGLTDLRQHI 420

361 CKTDPAPACGCEIIKILPHILDPKLLTAEDYSALRERVLVKLERMGLTLRQHIVT 420  
421 BEYWTPLDIOAKYYSNCGSIYGVVADRFKNGFKAPQORSSELSNLYFVGSVNPGGMPM 480  
422 BEYWTPLDIOAKYYSNCGSIYGVVADRFKNGFKAPQORSSELSNLYFVGSVNPGGMPM 480  
481 VTLSGQVRDKIVADLQ 497  
481 VTLSGQVRDKIVADLQ 497

## RESULT 3

JB0333  
AAU80333 standard; protein; 497 AA.

## AAU80333;

15-JUL-2002 (first entry)

Methylomonas 16a ORF9 crtN2 protein sequence.

Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;  
keratenoid; pigment; flavour; fragrance; open reading frame 9; ORF9;  
crtN2; crtN copy2; diapophytoene dehydrogenase enzyme.

Methylomonas sp.

WO2000220733-A2.

14-MAR-2002.

29-AUG-2001; 2001WO-US026852.

01-SEP-2000; 2000US-0229907P.

(DUPO ) DU PONT DE NEMOURS & CO E I.

Cheng Q, Koffas M, Norton KC, Odum JM, Picataggio SK;  
Rouviere PE, Schenzle A, Tomb J;

WPI; 2002-383051/41.

N-PSDB; ABK50089.

Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,  
isolated from Methylomonas 16a, useful for the production of isoprenoid  
compounds.

Claim 4; Page 81-82; 84pp; English.

The present invention relates to a new nucleic acid molecule encoding an  
isoprenoid biosynthetic enzyme isolated from Methylomonas 16a. The  
invention is useful for obtaining a nucleic acid molecule encoding an  
isoprenoid compound biosynthetic enzyme, and for the microbial production  
of isoprenoid compounds. The molecules of the invention are also useful  
for regulating isoprenoid biosynthesis in an organism and for producing  
recombinant organisms for producing various isoprenoid compounds. The  
nucleic acid is also useful for feed additive, for the production of  
keratenoids and their derivatives, isoprenoid intermediates, and as pure  
products useful as pigments, flavours and fragrances. The present amino  
acid sequence represents the Methylomonas 16a open reading frame 9 (ORF9)  
crtN2 (crtN copy2) copy 2 of diapophytoene dehydrogenase enzyme) protein  
of the invention, as described above

Sequence 497 AA;

Query Match 100.0%; Score 2598; DB 5; Length 497;

Best Local Similarity 100.0%; Pred. No. 1.5e-241;

Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNSNDNQRVIVIGAGLGLSLAISLATAGFSVQLIERKNDKVGKLNMTKDGFTDLGSPS 60

1 MNSNDNQRVIVIGAGLGLSLAISLATAGFSVQLIERKNDKVGKLNMTKDGFTDLGSPS 60

QY 61 ILTWPHIPEALFTGAGKNMADYVOIQKVEPHWRNFFEDGSGVIDLCEDAETORRELDKLG 120  
DB 61 ILTWPHIPEALFTGAGKNMADYVOIQKVEPHWRNFFEDGSGVIDLCEDAETORRELDKLG 120  
QY 121 GTYAQORFLDYSKNLCTETETAGYFAKGLDGFWDLLKFGYPLRSLLSFDVFRSMDQGVRR 180  
DB 121 GTYAQORFLDYSKNLCTETETAGYFAKGLDGFWDLLKFGYPLRSLLSFDVFRSMDQGVRR 180  
QY 181 FISDPKLVIELNYFIKYVGSSPYDAPALMNLPIQVHYGLWYVKGGMGMAQAMEKLAV 240  
DB 181 FISDPKLVIELNYFIKYVGSSPYDAPALMNLPIQVHYGLWYVKGGMGMAQAMEKLAV 240  
QY 241 ELGVEIRLDAEVSEIQKDGGRACAVKLANGDVLPADIVSNMVEVIPAMEKLLRSPASELK 300  
DB 241 ELGVEIRLDAEVSEIQKDGGRACAVKLANGDVLPADIVSNMVEVIPAMEKLLRSPASELK 300  
QY 301 KMORPEPSCSGLVLHLGVDRILYPOLAHNPFYSDPREHFDVAFKSHRLSDDPTIYLAVP 360  
DB 301 KMORPEPSCSGLVLHLGVDRILYPOLAHNPFYSDPREHFDVAFKSHRLSDDPTIYLAVP 360  
QY 361 CKTDPAPACGCEIIKILPHILDPKLLTAEDYSALRERVLVKLERMGLTLRQHIVT 420  
DB 361 CKTDPAPACGCEIIKILPHILDPKLLTAEDYSALRERVLVKLERMGLTLRQHIVT 420  
QY 421 BEYWTPLDIOAKYYSNCGSIYGVVADRFKNGFKAPQORSSELSNLYFVGSVNPGGMPM 480  
DB 421 BEYWTPLDIOAKYYSNCGSIYGVVADRFKNGFKAPQORSSELSNLYFVGSVNPGGMPM 480  
QY 481 VTLSGQVRDKIVADLQ 497  
DB 481 VTLSGQVRDKIVADLQ 497

## RESULT 4

ADA14538  
ID ADA14538 standard; protein; 497 AA.

XX AC ADA14538;

XX DT 06-NOV-2003 (first entry)

XX DE Methylomonas sp. 16a CrtN2 protein SEQ ID NO:8.

XX KW carotenoid biosynthetic enzyme; CBE;

XX KW omega-aldehyde-functionalised carotenoid;

XX KW omega-carboxy-functionalised carotenoid; C30 carotenoid;

XX KW carotenoid biosynthesis; antioxidant; microbial; pigment; feed additive;

XX KW CrtN2.

XX OS Methylomonas sp.

XX PN WO2003068917-A2.

XX PD 21-AUG-2003.

XX PF 11-FEB-2003; 2003WO-US004150.

XX PR 11-FEB-2002; 2002US-0355939P.

XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.

XX PI Cheng Q, Norton KC, Tao L;

XX DR WPI; 2003-697524/66.

XX DR N-PSDB; ADA14537.

XX FT New nucleic acid encoding carotenoid biosynthesis enzymes, useful for  
preparing functionalized carotenoids, e.g. pigments and antioxidants,  
PT also encoded polypeptides.

XX PS Claim 19; Page 99-102; 125pp; English.

XX XX

The present invention describes nucleic acid sequences encoding carotenoid biosynthetic enzymes (CBEs). Also described: (1) a CBE polypeptides; (2) a chimeric gene (CG) containing a CBE linked to regulatory sequences; (3) host cell transformed with CG; (4) obtaining a nucleic acid that encodes a CBE; (5) products of method (4); (6) producing an omega-aldehyde-functionalised carotenoid (X); (7) producing an omega-carboxy-functionalised carotenoid (Y); (8) producing C30 carotenoids (Z); and (9) regulating carotenoid biosynthesis in an organism. CBE sequences have antioxidant activity. CBEs can be used for microbial production of aldehyde- or carboxy-functionalised carotenoids and C30 carotenoids, and are potentially useful as pigments and antioxidants, e.g. as feed additives. The present sequence represents a CrtN2 CBE protein isolated from *Methylobacterium* sp. strain 16a, which is used in the exemplification of the present invention.

Query Match 100.0%; Score 2598; DB 6; Length 497;

Best Local Similarity 100.0%; Pred. No. 1.5e-241;

Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNSNDNRVTVIGAGLGLSAAISLATAGSVQLIEKNDKVGKLNIMTKDGTEDLGP 60

1 MNSNDNRVTVIGAGLGLSAAISLATAGSVQLIEKNDKVGKLNIMTKDGTEDLGP 60

61 ILTWPHFEALFTGAGNMDYVQIOKVEPHRNFFEDGSVIDLCEDAETORRELDKLP 120

61 ILTWPHFEALFTGAGNMDYVQIOKVEPHRNFFEDGSVIDLCEDAETORRELDKLP 120

121 GTVAQFORFLDYSKNLCTETAGYFAGKLDGFWDLLKPYGRLSLLSPDFVFRSMQGVRR 180

121 GTVAQFORFLDYSKNLCTETAGYFAGKLDGFWDLLKPYGRLSLLSPDFVFRSMQGVRR 180

181 FISDPKLVLEILNPKVGGSSPDAPALNNLLPYIOHYGLWTVKGMQMAQAEKLV 240

181 FISDPKLVLEILNPKVGGSSPDAPALNNLLPYIOHYGLWTVKGMQMAQAEKLV 240

241 ELGVEIRLDAEVSIIQKDGACAVKLANGDVLPAIVVSNMEVIPAMEKLLRSASELK 300

241 ELGVEIRLDAEVSIIQKDGACAVKLANGDVLPAIVVSNMEVIPAMEKLLRSASELK 300

301 KMQRFESCSGLVHLGVDRILYPOLAHNFFYSDHPREHFDVAFKSHRSLSDPTIYLVAP 360

301 KMQRFESCSGLVHLGVDRILYPOLAHNFFYSDHPREHFDVAFKSHRSLSDPTIYLVAP 360

361 CKTDPAQAPACCEIIKILPHIPLDPPKLLTAEDYSALRERVLVLERMGLTDLRQHIYV 420

361 CKTDPAQAPACCEIIKILPHIPLDPPKLLTAEDYSALRERVLVLERMGLTDLRQHIYV 420

421 BEYWTPLDIOAKYYSNQGSIYGVVADRFKLGKAPQSRSELSNLYFVGSGVNPFGGMPM 480

421 BEYWTPLDIOAKYYSNQGSIYGVVADRFKLGKAPQSRSELSNLYFVGSGVNPFGGMPM 480

481 VTLGSQLVRDKIVADLQ 497

481 VTLGSQLVRDKIVADLQ 497

RESULT 5

ID ADAL4544

KX ADAL4544 standard; protein; 497 AA.

KX ADAL4544;

KX 06-NOV-2003 (first entry)

KX Staphylococcus aureus CrtN2 protein SEQ ID NO.14.

KX carotenoid biosynthetic enzyme; CBE;

KX omega-aldehyde-functionalised carotenoid;

KX omega-carboxy-functionalised carotenoid; C30 carotenoid;

KX carotenoid biosynthesis; antioxidant; microbial; pigment; feed additive;

KX CrtN2.

XX Staphylococcus aureus.

XX WC2003068917-A2.

XX 21-AUG-2003.

XX 11-FEB-2003; 2003WO-US004150.

XX 11-FEB-2002; 2002US-0355939P.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Cheng Q, Norton KC, Tao L;

XX WPI; 2003-697524/66.

XX N-PSDB; ADA14543.

XX New nucleic acid encoding carotenoid biosynthesis enzymes, useful for preparing functionalized carotenoids, e.g. pigments and antioxidants, also encoded polypeptides.

XX Claim 19; Page 114-117; 125pp; English.

XX The present invention describes nucleic acid sequences encoding carotenoid biosynthetic enzymes (CBEs). Also described: (1) a CBE polypeptides; (2) a chimeric gene (CG) containing a CBE linked to regulatory sequences; (3) host cell transformed with CG; (4) obtaining a nucleic acid that encodes a CBE; (5) products of method (4); (6) producing an omega-aldehyde-functionalised carotenoid (X); (7) producing an omega-carboxy-functionalised carotenoid (Y); (8) producing C30 carotenoids (Z); and (9) regulating carotenoid biosynthesis in an organism. CBE sequences have antioxidant activity. CBEs can be used for microbial production of aldehyde- or carboxy-functionalised carotenoids and C30 carotenoids, and are potentially useful as pigments and antioxidants, e.g. as feed additives. The present sequence represents a CrtN2 CBE protein isolated from *Staphylococcus aureus*, which is used in the exemplification of the present invention.

XX Sequence 497 AA;

XX Query Match 49.6%; Score 1289; DB 6; Length 497;

XX Best Local Similarity 51.9%; Pred. No. 3.7e-115;

XX Matches 254; Conservative 85; Mismatches 148; Indels 2; Gaps 2;

QY 7 QRVTVIGAGLGLSAAISLATAGSVQLIEKNDKVGKLNIMTKDGTEDLGPSTLMPH 66

Db 3 KHIIVIGAGLGLSAAIRMAQSGYSVSLYEONKHGKVNHRHSDGFGDLGSIITMPY 62

QY 67 IFEALFTGAGNMDYVQIOKVEPHRNFFEDGSVIDLCEDAETORRELDKLGPGTYAQF 126

Db 63 IPEKLFYSKQMSDYVTKRLPHQWRSFFPDGTTIDLYEGIKETGQHAILSKQDIBEL 122

QY 127 QRFLDYSKNLCTETAGYFAGKLDGFWDLLKPYGRLSLLSPDFVFRSMQGVRRFISDPK 186

Db 123 QNVLNTRRIDRITEKGYFNYGLDLSQIIKFHGPLNALLINYVHTMQQAIDKRSNPF 182

QY 187 LVEILNLYFIVKVGSSPDAPALNNLLPYIOHYGLWTVKGMQMAQAEKLVAGVEI 246

Db 183 LRQMLGTFPIKVGSSSDAPAVLSMLFHMQOGLWYVGGIIHILANALEKLABEGVTI 242

QY 247 RLDAEVSBIQKDGACAVKLANGDVLPAIVVSNMEVIPAMEKLLRSASELKQKQF-F 305

Db 243 HTGARVDNIKTYQSRVGVRLDTGCFVKADYIISNNEVIPTKYKYLHLDTQRLKLEREF 302

QY 306 EPSCSGLVHLGVDRILYPOLAHNFFYSDHPREHFDVAFKSHRSLSDPTIYLVAPCKTDP 365

Db 303 EPASSGYVMHLGVACQYVLPOLAHNFFFTENAYNLNQVVFHEKVLPPDPTIYLVNTKTDH 362

QY 366 AQAPACCEIIKILPHIPLDPPKLLTAEDYSALRERVLVLERMGLTDLRQHIYVTEYWT 425

Db 363 TCAPVGYENIKVLPHIPIQ-DQPTTEDYAKFRDKILDKLKMGLTDLRKHIIYEDVMT 421

426 PLDIQAKYNSQSIYGVVADRFKXGLGFKAPQRSSELSNLYFVGGSVNPGGSMFMTLSG 485  
422 PEDIEKTRSRNGALYGVVADKXKNGKFKPKESQYFENLYFVGGSVNPGGSMFMTLSG 481  
486 QLVDRKIVA 494  
482 QQVADKINA 490  
RESULT 6  
ABM72616 standard; protein; 500 AA.  
ABM72616;  
20-NOV-2003 (first entry)  
Staphylococcus aureus protein #1856.  
Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;  
enzymatic assay; antibiotic target.  
Staphylococcus aureus.  
WO200294863-A2.  
28-NOV-2002.  
27-MAR-2002; 2002WO-IB002637.  
27-MAR-2001; 2001GB-00007661.  
(CHIR-) CHIRON SPA.  
Masignani V, Mora M, Scarselli M;  
WPI: 2003-120786/11.  
N-PSDB; AC74176.  
New Staphylococcus aureus protein, useful as a vaccine for treating or  
preventing Staphylococcal infection, specifically an infection caused by  
S. aureus, e.g. sepsis.  
Claim 1; SEQ ID NO 3712; 49pp; English.  
The invention relates to novel genes and encoded proteins from  
Staphylococcus aureus. A composition comprising the S. aureus protein, a  
nucleic acid encoding the protein, or an antibody to the protein, is  
useful as a pharmaceutical, particularly as a vaccine for treating or  
preventing infection due to Staphylococcus bacteria, specifically an  
infection caused by S. aureus. The composition is particularly useful for  
treating or preventing sepsis in a patient. The composition can also be  
used for diagnostics. The protein is also used in an assay for enzymatic  
studies and as a target for antibiotics. This sequence represents one of  
the novel S. aureus proteins of the invention  
Sequence 500 AA;  
Query Match 49.6%; Score 1289; DB 6; Length 500;  
Best Local Similarity 51.9%; Pred. No. 3.8e-115;  
Matches 254; Conservative 85; Mismatches 148; Indels 2; Gaps 2;  
7 QRVIVIGAGLGLSAISLATAGFSVQLIERNDKVGKLNMTKDGTFDLGPSTLMPH 66  
6 KHIIVIGGLGISAIRMAQSGYSVLYEQNNHIGKVNFRHESDGFGLGPSILTMPY 65  
67 IPEALFTGAGKMDYVQIKVEPHRWFFRDEGSDVIDLCEDAETORELKLGPCTAOF 126  
66 IFEKLFYSKKQMSDYVTKRPLPHWRSFDPDGTIDLYEGIKETGQHNAILSKQDIEEL 125  
127 QRLDYSKMLCTETAGYFAGKLDGFDLGLKPYGLRSLISFDVFRSMDQGVRRFISDPK 186  
126 QYLYNTRIDRITEKGFVNGLOTLISQIIRKHPFLNALINYDVHTMQAIDKXISNPY 185

187 LVEILNLYPIKTVGSSPYDAPALMNLPIYQYEGYGLWTVYKGMGMAQAMEKLAVELGVEI 246  
186 LRQMLGYPIKTVGSSSYDAPAVLSNLFHMQEQGLWYVEGGIHHLANALEKLAREEGVTI 245  
247 RLDAEVSIOKODGRACAVKLANGVLPADIVSNMEVI PAMEKILRSPASELYKMQR-F 305  
246 HTGARVDNIKTYQRRVTGVRDGTGEFVKADYIISNMEVIPTKYLIHLDTQRLNKLREFF 305  
306 EPSCGLVHLGVDRLYPQLAHNFFYSDHPREHFDVAFKSHRSLSDPTIYLVAPCKTDP 365  
306 EPASSGYWHLGAVACQYFQLAHNFFFTENALYNYQVFEKVLDPDPTIYLVVNTKTDH 365  
366 AQAPAGCEIILPHIPHLDPDKLLTAEDYSALSERVLVRLMGLTDLRQHIIVTEEYWT 425  
366 TQAPVGENIKVLPHIPYIQ-DQPTTEDIYAKFRDKILDKLEKGLTDLRKHIIYEDVWT 424  
426 PEDIQAKYNSQSIYGVVADRFKXGLGFKAPQRSSELSNLYFVGGSVNPGGSMFMTLSG 485  
425 PEDIEKTRSRNGALYGVVADKXKNGKFKPKESQYFENLYFVGGSVNPGGSMFMTLSG 484  
486 QLVDRKIVA 494  
485 QQVADKINA 493  
RESULT 7  
ABU16031  
ID ABU16031 standard; protein; 497 AA.  
XX AC ABU16031;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #1558.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Staphylococcus aureus.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR WPI: 2003-029926/02.  
XX DR N-PSDB; ACA19901.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 25; SEQ ID NO 43955; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: the sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 CC Sequence 497 AA;

Query Match 49.5%; Score 1287; DB 6; Length 497;  
 Best Local Similarity 51.9%; Pred. No. 5.8e-115;  
 Matches 254; Conservative 84; Mismatches 149; Indels 2; Gaps 2;  
 2y 7 ORVIVIGAGLGLSAAISATAGSVOLEIKNDKVGKGLNMTXGDFGLGSPSILTMPH 66  
 Db 3 KHLIVIGGLGISAIRWAGSIVSVIEQTHIGKVNHEHSDGFDGLSPSILTMPI 62  
 Qy 67 IFEALFTGAGNMADYVOIQKVEPHRNFTEDGSDVIDCEDAETQRELDKLGPTVAQF 136  
 Db 63 IFEKLFYSKXQMSDYVTKRLPHQWRSFPDGTGTTIDLYEGIKETGQFNAILSKQDIEEL 122  
 Qy 127 QFELYSNKLTETAGYFAKGLDGFWDLLAFYGLPLSLLSFDVFRSMDOGVRRFTSDPK 186  
 Db 123 QYLNWTRIDRITKGFYVNGVGLDTLSQIIKPHGLPLALINDYVHTWQQAIDKRISNPY 182  
 Qy 187 LVEILNFIKYVGSFYDAPALMNLPIQYHGLWYVKGMYGMAQAKELVGLGVEI 246  
 Db 183 LQMLGYFIKYVGSFYDAPAVLSMLPMQOEGQLWYVEGGIHLNALAKELAREGVTI 242  
 Qy 247 RLDAVSBIGKQDGCACAVKLANGVLPADIVSNMVEVTPAMEKLLSPASELKMQR-F 305  
 Db 243 HFGARVDNIXYQRRVTGVRDTGFEFVADYIISNMEVITPYKYLIIHLDTQRLNKLREF 302  
 Qy 306 EPSCSLVHLGVDRLYPQLAHNFFYGDHPRHFDVAFKSHRLSDDPITYLVAPCKTDP 365  
 Db 303 EPASSGVWELGVACQYQLAHNFFFTENAYLNYQVFEHEKVLDPDPTIYLVNTKTDH 362  
 Qy 366 AAPAGCEIILKILPHIPLDLPKLLTADYSALRERVLVLERGLTDLRQHIYTESYWT 425  
 Db 363 TOAPGYENIKVLPHPYIQ-DQPTTEDYAKFRKLDKLEKGLTDLARKHIYEDVWT 421  
 Qy 426 PLDIOAKYVNSGSIYGVVADRFKNLGFAPQSRSELNLYFVGGSVNPGGMPVTLTG 485  
 Db 422 PEDIEKVRNSGAIYGVVADKKNKGFKEQYFENLYFVGGSVNPGGMPVTLTG 481  
 Qy 486 QIVRDKIVA 494  
 Db 482 QQVADKINA 490

RESULT 8  
 ABU43877  
 ID ABU43877 standard; protein; 502 AA.  
 XX  
 AC ABU43877;

XX 19-JUN-2003 (first entry)  
 DT Protein encoded by Prokaryotic essential gene #29404.  
 DB Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 KW Staphylococcus haemolyticus.  
 XX WO200277183-A2.  
 XX 03-OCT-2002.  
 PD 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI: 2003-029926/02.  
 XX N-PSDB; ACA47747.  
 DR New antisense nucleic acids, useful for identifying proteins or screening  
 XX for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PT  
 XX Claim 25; SEQ ID NO 71801; 1766pp; English.  
 PS The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 CC  
 CC Sequence 502 AA;

Query Match 49.3%; Score 1280; DB 6; Length 502;  
 Best Local Similarity 52.1%; Pred. No. 2.8e-114;  
 Matches 255; Conservative 80; Mismatches 152; Indels 2; Gaps 2;



XX	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW	detection; food; gene expression; plant; animal; microorganism; toxin;
KW	antibiotic; biopesticide; virulence factor; disease model; plague;
KW	whooping cough.
OS	Photorhabdus luminescens.
XX	
XX	WO200294867-A2.
PN	
XX	
PD	28-NOV-2002.
XX	
PD	07-FEB-2002; 2002WO-IB003040.
XX	
PF	07-FEB-2001; 2001FR-00001659.
XX	
PR	(INSP ) INST PASTEUR.
PA	(CNRS ) CNRS CENT NAT RECH SCI.
XX	
XX	Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI	Buchrieser C;
PI	
XX	
DR	WPI; 2003-148459/14.
XX	
XX	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT	useful e.g. as therapeutic antimicrobials and agricultural pesticides.
PT	
PS	Claim 2; SEQ ID NO 3220; 1205pp; French.
XX	
XX	The invention relates to the isolation of genes and their encoded
CC	proteins from Photorhabdus luminescens. The isolated sequences are
CC	sources of probes and primers for detecting the genome of P. luminescens
CC	and related species; to study polymorphisms; for gene analysis and for
CC	detection/amplification of the genes. Antibodies (Ab) raised against the
CC	polypeptides encoded by the genes are used for detection/identification
CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC	carry a gene-containing vector are used to select compounds that
CC	modulate, regulate, induce or inhibit expression of the genes in plants,
CC	animals or microorganisms other than P. luminescens and are able to alter
CC	response or sensitivity to toxins and antibiotics produced by P.
CC	luminescens. Cells transformed to express the genes are useful for
CC	recombinant production of the proteins, particularly toxins and
CC	antibacterials useful as insecticides, bactericides and fungicides. The
CC	genes, proteins, vectors containing the genes and Ab are also useful
CC	therapeutically (to treat microbial infection by bacteria or fungi that
CC	are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC	biopesticides. Other uses of the genes and the proteins are as virulence
CC	factors and for identifying targets of human diseases for which P.
CC	luminescens is a model (particularly plague and whooping cough). This
CC	sequence represents one of the isolated P. luminescens proteins
XX	
XX	Sequence 494 AA;
XX	
XX	Query Match 27.4%; Score 713; DB 6; Length 494;
XX	Best Local Similarity 32.1%; Pred. No. 1.6e-59;
XX	Matches 158; Conservative 97; Mismatches 231; Indels 6; Gaps 3
QY	8 RVIVTICAGLGLSMAISLATAGFSVQLIEKNKDVGKGLNMTKDGFTFDLGRSLITMDEHI 67
DB	3 KALVTCAGFGGLAIRLQISAGITCTILEQRDKGGRAYVYKGGFTFDAGPVTAPNV 62
QY	68 FEALPTGAGNMADYVQIQKVEPHRWPFDDGSGVIDLCEDAETQRELBKLGPGTYAQPO 127
DB	63 IEELFTQAGKWMADYVDLPLVPHYFRLCWESGKRMFDYNDQOHLAQIHFTFPRDVGNGYR 122
QY	128 RLDYSXNLCCTETEAGYPAKGLDGFWDLLKPYGLRSLISFDVFRSMDQGVRRFISDPKL 187
DB	123 RLDYSREAFNE--GYLKLGTVPFLSPRDLMSAQPQLIRLHAWRSVYSQVAFIKDEL 179
QY	188 VEILNYFIKYVGSSYPDALNLLPYIQVHYGLWYKGYGMQAMKELAVEIGVEIR 247
DB	180 ROAFSGHLLIGNPFAASSYITLHALERWGWFVPRGTSALVAMVKLFTDGGEIE 239

QY	248	LDAREVSIQKQDGRACACAKLANGDVLPA	DIIVSNMEVI	PAMEKLL-RSPA--SELKKMOR	300
DB	240	LAARVKSFTTHGVRVTCVGLADGRNTC	CDVASNADV	VIHTYKELLCQHPGFIARARTLVR	299
QY	305	FEPSCGLVLHLGVDRILYPOLAHENPFY	SDHPREHDAVFKSHR	LSDDPTIYLVAPCKTD	364
DB	300	KRMNSLFLVYFGLNHHTQLAHTTVC	FGFRYKELIEDIFYH	DRLSDEDSFLYLHAPSVD	359
QY	365	PAQAPAGEIILKILPHILPDPDKLLTA	EDYSALRERVLVKLERMGLTDLRQ	HIVTEEYV	424
DB	360	PSLAPRCASYVLAVPPLHGTANLNWD	IEGPELRDRIEAYLEKYMPGLLQ	LVVHRIF	419
QY	425	TPLDIQAKYVSNQSGISYGVVADRFK	NLGFKAPOSSELSNLIFYVGS	VNPGGMPWTLUS	484
DB	420	TPFDFRDQLNAHLGSAFSEFPELLT	QSAWFRPHNRDNRI	DNLYLVGAGTHPGAGIPGVIGS	479
QY	485	GQLVRDKIVADL	496		
DB	480	AKATATLMLEDI	491		
XX	RESULT 11				
XX	AAR13985				
XX	ID	AAR13985	standard; protein; 489 AA.		
XX	AC	AAR13985;			
XX	XX	24-OCT-2003	(revised)		
XX	XX	25-MAR-2003	(revised)		
XX	XX	26-NOV-1991	(first entry)		
XX	XX	Phytoene dehydrogenase-4H	from PARC146D.		
XX	XX	GGPP; carotenoid; phytoene; zeaxanthin; lycopene.			
XX	XX	Pantoea agglomerans; EHO-10	(E. vulneris - ATCC 39368).		
XX	XX	WC9113078-A.			
XX	XX	05-SEP-1991.			
XX	XX	04-MAR-1991;	91WO-US001458.		
XX	XX	02-MAR-1990;	90US-00487613.		
XX	XX	18-MAY-1990;	90US-00525551.		
XX	XX	03-AUG-1990;	90US-00562674.		
XX	XX	28-FEB-1991;	91US-00662321.		
XX	XX	(STAD )	AMOCO CORP.		
XX	XX	Ausich RL, Brinkhaus FL, Mukharji I, Proffitt JH, Yarger JG;			
XX	XX	Yen HC;			
XX	XX	WPI; 1991-281410/38.			
XX	XX	N-PSDB; AAQ13720.			
XX	XX	Biosynthesis of carotenoid(s) in genetically engineered hosts - using DNA			
XX	XX	encoding enzymes from Erwinia herbicola.			
XX	XX	Disclosure; Fig 15(1-4); 313pp; English.			
XX	XX	There are a total of six relevant genes in a 7900 bp region that cause E.			
XX	XX	coli cells to produce GGPP and the carotenoids phytoene through			
XX	XX	zeaxanthin diglucoside, which is the final prod. identified in the			
XX	XX	carotenoid pathway contd. in plasmid pARC376 (contg. a ca. 13 kb			
XX	XX	chromosomal DNA fragment isolated by Perry et al., J. Bacteriol., 168: 607			
XX	XX	(1986). The genes for geranylgeranyl pyrophosphate (GGPP) synthase,			
XX	XX	phytoene synthase, phytoene dehydrogenase-4H, lycopene cyclase, beta-			
XX	XX	carotene hydroxylase, and zeaxanthin glycosylase are represented in			
XX	XX	AAQ13716, AAQ13718, AAQ13719, AAQ13722, AAQ13724 and AAQ13726			
XX	XX	respectively. Recombinant expression plasmids can be used to produce			
XX	XX	large amts. of the enzymes and hence large amts. of the carotenoids which			
XX	XX	they synthesize. Carotenoids are pigments with a variety of applications.			





AAW01123 standard; protein; 489 AA.  
AAW01123;  
16-OCT-2003 (revised)  
25-MAR-2003 (revised)  
11-DEC-1996 (first entry)  
Phytoene dehydrogenase-4H encoded on pARC146D.  
GGPP synthase; lycopene cyclase; phytoene synthase; beta-carotene;  
phytoene dehydrogenase-4H; carotenoid; C40; zeaxanthin; diglucoside;  
pigment; food colourant; chloroplast transit peptide; increase yield;  
tobacco ribulose bis-phosphate carboxylase-oxygenase.  
Pantoea agglomerans.  
US5530188-A.  
25-JUN-1996.  
21-JUL-1993; 93US-00095726.  
02-MAR-1990; 90US-00487613.  
18-MAY-1990; 90US-00525551.  
03-AUG-1990; 90US-00562674.  
28-FEB-1991; 91US-00662921.  
30-OCT-1991; 91US-00785566.  
(STAD ) AMOCO CORP.  
Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;  
Ausich RL;  
WPI; 1996-308823/31.  
N-PSDB; AAT40793.  
Increasing prodn. of total carotenoid(s) in a higher plant - by  
transforming with vector encoding chloroplast transit peptide operably  
linked to the Erwinia herbicola lycopene cyclase structural gene.  
Example 12; Col 97-100; 93pp; English.  
The present sequence is that of a recombinant Erwinia herbicola phytoene  
dehydrogenase-4H (from pARC146D) which produces lycopene biosynthetically  
from phytoene through four sequential dehydrogenation reactions. Other  
enzymes involved in the carotenoid biosynthesis pathway include  
geranylgeranyl pyrophosphate (AAW01119), phytoene synthase (W01121) and  
lycopene cyclase (AAW01125). Manipulation of, in partic., lycopene  
cyclase gene (AAT40795), by in frame linkage to the chloroplast transit  
peptide (AAW01124) of the tobacco ribulose bis-phosphate carboxylase-  
oxygenase gene can lead to increased production of total carotenoids in  
the chloroplast of transformed plants as compared to native, non-  
transformed plants of the same type. Beta-carotene is an effective and  
apparently harmless food colourant and is also in the pathway for  
biological synthesis of further C40 carotenoids such as zeaxanthin and  
zeaxanthin diglucoside. (Updated on 25-MAR-2003 to correct PF field.)  
(Updated on 16-OCT-2003 to standardise OS field)  
Sequence 489 AA;

QY 127 QRFLDYSKNLCTETETAGYFAKG---LDGEWDLLKFKYGLRSLLSFDVFRSMQGVRRFIS 183  
DB 122 RRFLAYSOAVPQE---GYLFLGSPVFLSPDKMLR-AGP---QLLKLQAWQSVTVQSVSRFIE 175  
QY 184 DPKLVEILNTFYIKYVGSSPYDAPALANLLPYIQYHYCLANVYKGYGMAQAKELAVELG 243  
DB 176 DEHLRQAFSPHSLLVGNNPFTTSSYYTLIHALERBWGVWFPPEGGTGLVNGWVKLFDTLGG 235  
QY 244 VEIRLDAVSEIKQDGRACAVKLANGDVLPAIIVVSNMEVTPAKMKLLRS-PASBLK-- 300  
DB 236 GEIELNARVELLVADNRVSQVELADGRIFPTDAVASNADVVNTYKKLLGTIPVGOKRAA 295  
QY 301 KMORFEPSCGLVHLGVDRLYQOLAHENFFSDHPRHFDVAFKSHRLSDOPTIYLVAP 360  
DB 296 RLERKSMNSLFLVLYTGLNQPHSQLAHTICFCGPRYRELIDEFTGSALADDFSLYHSP 355  
QY 361 CKTDPAQAPAGCEIILPHIPLHDPDKLITADYSALZERVLVLERMGLTDLRQHIVT 420  
DB 356 CVTDPSLAPPFCASFYVLAPVPHLGNAPLDWAGGPKLRDRIPDYLEERYMPCGLRSQVLT 415  
QY 421 EYWTPLDIQAKYYSNQSIIY-----GVVADRFKNLGFKAQORSSSELNLYFVGGSV 472  
DB 416 QRIFRQTSRPAWIAILGSLFIEPPLTQGLFA-----ANATRHSNLYLYAAGT 464  
QY 473 NPGGEMPMVTLGQLVRDKIVADIQ 497  
DB 465 HPGAGIFGVVGLAEASTASLWIEDLQ 489  
RESULT 14  
AAW00174  
ID AAW00174 standard; protein; 489 AA.  
XX AC AAW00174;  
XX DT 16-OCT-2003 (revised)  
XX DT 25-MAR-2003 (revised)  
XX DT 17-OCT-1996 (first entry)  
XX DE Phytoene dehydrogenase-4H encoded by DNA derived from pARC146D.  
XX KW E. herbicola; geranylgeranyl pyrophosphate synthase; pARC376;  
XX KW GGPP synthase; biosynthesis; carotenoid; lycopene;  
XX KW farnesyl pyrophosphate; phytoene; FPP; isopentenyl pyrophosphate; IPP;  
XX KW tail to tail dimerisation; phytoene synthase; phytoene dehydrogenase-4H;  
XX KW food colourant; herbicide; norflurazon.  
XX OS Pantoea agglomerans.  
XX PN US5530189-A.  
XX PD 25-JUN-1996.  
XX PF 22-JUL-1993; 93US-00096043.  
XX PR 02-MAR-1990; 90US-00487613.  
XX PR 18-MAY-1990; 90US-00525551.  
XX PR 03-AUG-1990; 90US-00562674.  
XX PR 28-FEB-1991; 91US-00662921.  
XX PR 30-OCT-1991; 91US-00785568.  
XX (STAD ) AMOCO CORP.  
XX PA Yen HB, Brinkhaus FL, Proffitt J, Mukharji I, Yarger J;  
XX PI Ausich RL;  
XX WPI; 1996-308824/31.  
XX DR N-PSDB; AAT37095.  
XX PT DNA encoding Erwinia herbicola phytoene dehydrogenase-4H - used for  
XX PT prodn. of lycopene, and to produce transgenic plants resistant to  
XX PT norflurazon.

Query Match 27.1%; Score 703.5; DB 2; Length 489;  
Best Local Similarity 32.9%; Pred. No. 1.3e-58;  
Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;  
2Y 7 QRVTIVIGLGLSNAISLATAGSVOLIKNDKVGKLMITKDGFTFDLGPSTLTMPEH 66  
2B 2 EKTWVGAGFGGLALRLQAAQIPTVLLQDRKPGGRAYVWHDQGGFTFDGFTVIDPT 61  
2Y 67 IFEALFTGAGKNADYVQIQKVEPHWRNFEDSGSVIDLCEDAETQRRELKLGPGTYAQF 126  
2B 62 ALREALFTLAGRMEDYVRLLPVKFFYLKWSGKTLVDYANDSPLEAQITQFNPRDVEGY 121

Example 4; Fig 15; 87pp; English.

This sequence represents Erwinia herbicola phytoene dehydrogenase encoded by DNA from the plasmid pARC146D. Phytoene dehydrogenase is an enzyme which is involved in the biosynthesis of carotenoids, esp. lycopene, from the ubiquitous precursor, farnesyl pyrophosphate. In E. herbicola, phytoene has been found to be formed biosynthetically in a two-step process. The initial step is the condensation of farnesyl pyrophosphate (FPP) and isopentenyl pyrophosphate (IPP) to form GGPP. This reaction is catalysed by GGPP synthase. This first step is immediately followed by a tail to tail dimerisation of GGPP, catalysed by the enzyme phytoene synthase, to form phytoene. Lycopene is produced from phytoene by the catalytic action of phytoene dehydrogenase-4H. The genes encoding components of the lycopene biosynthesis pathway may be used to transform a host cell for the commercial production of lycopene which is used as a food colourant. Plants transformed with the phytoene dehydrogenase-4H coding sequence are protected from the herbicide norflurazon. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 16-OCT-2003 to standardise OS field.)

Sequence 489 AA;

Query Match 27.1%; Score 703.5; DB 2; Length 489;  
Best Local Similarity 32.9%; Pred. NO. 1.3e-58;  
Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;  
7 QRVVIVGAGLGLSAAISLATAGFSVQLIEKNKDKVGGKLNIMTKDGTFFDLGSPILTMPH 66  
2 EKTIVVIGAGFGGLALAIQLQAGITVLLQKDKPGRAYVVMHQGTFFDAGTVIDTPT 61  
67 IFEALFTGAGKNMADYVQIQVKEPHRNFFEDSGVIDLCEDAETQRRLDKLPGTYAQF 126  
62 ALLEALFTLAGRMEDYVRLLPVKPFYRLCWESGKTLDYANDSFLEAQITQFNPRDVEGY 121  
127 QRFLLYSKNLCTETETAGYFAGK---LDGFWLLKPYGFLRSLLSFDVFRSMDQGVRRFIS 183  
122 RFLAYSQAVFOE---GYLRGVSVPFLSFRDMLR-AGP---QLLKQAWQSVYQSVRFIE 175  
184 DPKLVEILNYFIKYVSSPDYAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELG 243  
176 DEHLRQAFSHLLVGGNPFPTSSYITLIHALERWGVWFFEGGFGALVNGMKVLFDTLG 235  
244 VEIRLDAEYSEIQKQGRACAVKLANGDVLPAIVVSNMEVIPAMEKILRS-PASELK-- 300  
236 GEIELNARVEELVWADNRVSVQLADGRIFDTDAVASNADVVNTYKLLGTIPVGQKRAA 295  
301 KMQRFEPSCSGLVHLGVDRLYVQLAHNFFYSQDHPREHFDVAFKSHRLSDPTIYLAP 360  
296 RLKRSMSNSFLVYFGLNQPHSQLAHTICFGPRYRELIDEIFTGSALADDFSLYLHSP 355  
361 CKTDPAQAGCEIILKILPHILPDKLLTAEDYSALRERVLKLERMGLTDLRQHVT 420  
356 CVTDFELAPPCCASTVTLAPVHLGNAPLDWAQGPGLKURIRFDYLEERYMGLRSQVLT 415  
421 ERYWTPLDIQAKYNGQSIY-----GVVADRPNLGPAPQSRSELSNLYFVGGSV 472  
416 QRIFTRQTSRHAWIALGSLFTEPPSLTQGLFA-----ANATRHSNLYLVAAGT 464  
473 NFGGGMNMTLSGLVDRKIVADIQ 497  
465 HPGAGIGVGVGLAESTASLMIEDLQ 489

RESULT 15

AW32475

D AAW32475 standard; protein; 489 AA.

X AAW32475;

K AAW32475;

K 17-OCT-2003 (revised)

F 25-MAR-2003 (revised)

F 15-JAN-1998 (first entry)

Erwinia herbicola phytoene dehydrogenase-4H.  
Geranylgeranyl pyrophosphate synthase; Erwinia herbicola; GGPP;  
lycopene cyclase; recombinant enzyme; transgenic organism; beta-carotene;  
yeast; plant; vitamin A; cancer.

Pantoea agglomerans.

US5656472-A.

12-AUG-1997.

07-JUN-1995; 95US-00473512.

02-MAR-1990; 90US-00487613.

18-MAY-1990; 90US-00525551.

03-AUG-1990; 90US-00562674.

28-FEB-1991; 91US-00662921.

21-JUL-1993; 93US-00095726.

(STAD ) AMOCO CORP.

Mukharji I, Proffitt J, Ausich RL, Yarger J, Yen HB;

Brinkhaus FL;

XX WPI; 1997-414592/38.

DR N-PSDB; AAT91546.

DNA encoding Erwinia herbicola lycopene cyclase - for producing recombinant enzyme, and transgenic organisms with increased beta-carotene levels.

Dislosure; Fig 15; 102pp; English.

A novel DNA molecule has been isolated which encodes an Erwinia herbicola lycopene cyclase enzyme that converts lycopene to beta-carotene. The DNA molecule comprises at least 1125 bp and is present in the plasmids pARC147, pARC1509, pARC1510 and pARC1520. The present sequence represents the amino acid sequence corresponding to the structural gene for phytoene dehydrogenase-4H. The new DNA molecule can be used to produce the recombinant enzyme and transgenic organisms, e.g. Yeasts or Plants, with increased beta-carotene levels. Beta-carotene is used as a colourant in margarine and butter and as an intermediate for vitamin A, and may prevent cancer. (Updated on 25-MAR-2003 to correct PF field.) (Updated on 17-OCT-2003 to standardise OS field.)

Sequence 489 AA;

Query Match 27.1%; Score 703.5; DB 2; Length 489;

Best Local Similarity 32.9%; Pred. NO. 1.3e-58;

Matches 166; Conservative 96; Mismatches 212; Indels 31; Gaps 8;

7 QRVVIVGAGLGLSAAISLATAGFSVQLIEKNKDKVGGKLNIMTKDGTFFDLGSPILTMPH 66

2 EKTIVVIGAGFGGLALAIQLQAGITVLLQKDKPGRAYVVMHQGTFFDAGTVIDTPT 61

67 IFEALFTGAGKNMADYVQIQVKEPHRNFFEDSGVIDLCEDAETQRRLDKLPGTYAQF 126

62 ALLEALFTLAGRMEDYVRLLPVKPFYRLCWESGKTLDYANDSFLEAQITQFNPRDVEGY 121

127 QRFLLYSKNLCTETETAGYFAGK---LDGFWLLKPYGFLRSLLSFDVFRSMDQGVRRFIS 183

122 RFLAYSQAVFOE---GYLRGVSVPFLSFRDMLR-AGP---QLLKQAWQSVYQSVRFIE 175

184 DPKLVEILNYFIKYVSSPDYAPALMNLPIYQHYGLWYVKGMYGMAQAMEKLAVELG 243

176 DEHLRQAFSHLLVGGNPFPTSSYITLIHALERWGVWFFEGGFGALVNGMKVLFDTLG 235

244 VEIRLDAEYSEIQKQGRACAVKLANGDVLPAIVVSNMEVIPAMEKILRS-PASELK-- 300

236 GEIELNARVEELVWADNRVSVQLADGRIFDTDAVASNADVVNTYKLLGTIPVGQKRAA 295

301 KMQRFEPSCSGLVHLGVDRLYVQLAHNFFYSQDHPREHFDVAFKSHRLSDPTIYLAP 360

296 RLERKXNSLSFVLYFGLNQPHSOLAHTICGPRYRELIDEIFTGSALADDFSLYHSP 355  
361 CKTDPAQAPAGCEIILPHIPHLDPDKLLTAEDYSGALRERVVLKLERMGLTDLRQHIWT 420  
356 CVIDPSLAPPCASFVYLAAPVPHLGNAPLDWAGGPKLRDRIFDYLEERYMPLRSSQLVT 415  
421 ERYWTPLDIOAKYYSNOGSY-----GWVADRFNGLGFKAPQRSSELSNLYFVGGSV 472  
416 QRIFTRQTSRHWIAIIGSLFIEPPSLTQGLFA-----ANATRHSNLYLVAAGT 464  
473 NPGGWPVWTLGSLYRDKIVADLQ 497  
465 HPGAGIFGVVGLAESTASLMIEDLQ 489

Search completed: February 29, 2004, 14:43:47  
Job time : 60.8617 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:33:49 ; Search time 42.0179 Seconds  
(without alignments)  
3837.172 Million cell updates/sec

US-09-941-947a-22

fect score: 2/11  
quence: 1 MANTKHIIIVAGGGGLGAG.....KRVRFKDIHSAWLKAKA 511

oring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

- SPTREMBL 25:\*
- 1: sp archaea:\*
- 2: sp bacteria:\*
- 3: sp fungi:\*
- 4: sp human:\*
- 5: sp invertebrate:\*
- 6: sp mammal:\*
- 7: sp mhc:\*
- 8: sp organelle:\*
- 9: sp phase:\*
- 10: sp plant:\*
- 11: sp rodent:\*
- 12: sp virus:\*
- 13: sp vertebrate:\*
- 14: sp unclassified:\*
- 15: sp virus:\*
- 16: sp bacteriopl:\*
- 17: sp archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	1009	37.2	537	16 Q7UF87	Q7uf87 rhodopirell
2	929.5	34.3	490	16 Q8Y8A8	Q8y8a8 listeria mo
3	915.5	33.8	488	16 Q92D21	Q92d21 listeria in
4	883.5	32.6	517	2 QZG34	Q3zge4 heliobacill
5	808	29.8	498	16 Q88W3	Q88w3 lactobacill
6	794.5	29.3	499	16 Q8YKN8	Q8ykn8 anabaena sp
7	783.5	28.9	499	2 Q44318	Q44318 anabaena sp
8	739.5	27.3	511	16 Q8ENM0	Q8enm0 oceanobacil
9	729	26.9	514	2 Q8KZ28	Q8kz28 uncultured
10	726	26.8	513	2 Q840T3	Q840t3 rhodocyclu
11	725	26.7	511	2 Q3UP98	Q3up98 rhodocyclu
12	723	26.7	506	2 Q9KIX2	Q9kix2 bradyrhizob
13	713.5	26.3	492	2 Q47845	Q47845 pantoea agg
14	705.5	26.0	504	2 Q93C17	Q93c17 xanthobacte
15	703.5	25.9	492	2 Q8GCS1	Q8gcs1 pantoea ste
16	702.5	25.9	502	16 Q99R76	Q99r76 staphylococ

17	702.5	25.9	502	16 Q8NUQ6	Q8nuq6 staphylococ
18	699.5	25.8	506	2 Q07855	Q07855 staphylococ
19	683.5	25.2	492	2 Q8VUJ5	Q8vuj5 pantoea agg
20	673	24.8	582	3 Q13506	Q13506 phaffia rho
21	668	24.6	582	3 Q72858	Q72858 phaffia rho
22	651.5	24.0	548	16 Q9RW08	Q9rw08 deinococcus
23	650	24.0	579	3 Q8X020	Q8x020 gibberella
24	640.5	23.6	494	2 P94790	P94790 flavobacter
25	631	23.3	499	16 Q8XK44	Q8xk44 clostridium
26	628	23.2	579	3 Q9Y7H8	Q9y7h8 mucor circi
27	624.5	23.0	526	2 Q06757	Q06757 erythroba
28	610.5	22.5	501	2 Q9RLH4	Q9rlh4 paracoccus
29	596	22.0	514	17 Q27835	Q27835 methanobact
30	567	20.9	534	16 Q7UJ66	Q7uj66 rhodopirell
31	564.5	20.8	523	16 Q9R1X9	Q9rix9 streptomyc
32	552	20.4	499	2 Q9KS66	Q9ks66 mycobacteri
33	547.5	20.2	512	17 Q9HPD8	Q9hpd8 halobacteri
34	540	19.9	501	3 Q72857	Q72857 phaffia rho
35	531	19.6	588	2 Q83X01	Q83x01 streptomyc
36	521.5	19.2	488	2 Q7WT72	Q7wt72 marine bact
37	521	19.2	513	16 Q93HP2	Q93hp2 streptomyc
38	514	19.0	494	16 Q8ENM2	Q8enm2 oceanobacil
39	504	18.6	485	2 Q50893	Q50893 myxococcus
40	502.5	18.5	497	16 Q99R73	Q99r73 staphylococ
41	499.5	18.4	497	16 Q8NUQ3	Q8nuq3 staphylococ
42	485.5	17.9	536	17 Q9HP88	Q9hpb8 halobacteri
43	484	17.9	507	2 Q9EXL0	Q9exl0 streptomyc
44	483	17.8	549	2 Q91QX5	Q91qx5 corynebacte
45	482.5	17.8	544	2 Q8KP34	Q8kp34 agromyces m

ALIGNMENTS

RESULT 1

Q7UF87 PRELIMINARY; PRT; 537 AA.  
AC Q7UF87;  
DT 01-OCT-2003 (TREMELrel. 25, Created)  
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE Phytoene dehydrogenase (EC 1.14.99.-).  
GN CRT1 OR R810270.  
OS Rhodopirellula baltica.  
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
OC Planctomycetaceae; Pirellula.  
OX NCBI\_TaxID=117;  
RN 11;\_TaxID=117;  
RP SEQUENCE FROM N.A.  
RC STRAIN=1;  
RX MEDLINE=22735913; PubMed=12835416;  
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
RA Schlesner H., Amann R., Reinhardt R.;  
RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
RT strain 1".  
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
DR EMBL; EX294151; CAD78796.1; -;  
KW Oxidoreductase; Complete proteome.  
SQ SEQUENCE 537 AA; 60136 MW; A16C389C746949B6 CRC64;

Query Match 37.2%; Score 1009; DB 16; Length 537;

Best Local Similarity 41.5%; Pred. No. 1.3e+68;

Matches 204; Conservative 97; Mismatches 181; Indels 10; Gaps 6;

QY 5 KH:ITVAGPGGLCAGMLLSQRFKVSIFDKHAIIGRNRPINNGTFTGCPFFLLMKG 64

DB 25 KKVWVGAGPGGLASAMQLAAGGCDVTILERRGVGGRTSAIEDIDFRFCGPTFFLYPR 84

QY 65 VLDEMFEICERSDYELFLP---LSPWYRLY-DDRDIFVYSDRENNAELQRFVEGT 120

DB 85 VLDEIFHSTGR---DLMEVFMERLDPOYRLTFGGGGQLDCTPMDENDMDQIAQSPQDV 141



DR InterPro: IPR001100; Pyr redox.  
 DR Pfam: PF01593; Amino oxidase; 1.  
 DR PRINTS: PR00419; ADXRTASE.  
 DR PRINTS: PR00411; PNDRTASEI.  
 DR ProDom: PD139017; Phytin dehydro; 1.  
 SQ SEQUENCE 517 AA; 58340 MW; BB933977999C4587 CRC64;

Query Match 32.6%; Score 883.5; DB 2; Length 517;  
 Best Local Similarity 35.2%; Pred. No. 4.9e-59;  
 Matches 174; Conservative 117; Mismatches 194; Indels 9; Gaps 6;

QY	3	NKHHIIIVGAGPGGLCAGMLLSORGFKVSIIDKHAIIGGRNPINMGFTDTGTFLM	62
DB	7	NLGVIVVGSGAGAAVELANQGWDVTLEKEATPGGLSAIOAEGYSIDVGGTIMMM	66
QY	63	KGVLDMEFELCERSSDYLFPLSPWRLLYDDRDIFVYS-DRENMAEQLORVFDEGD	121
DB	67	NDVHQVFKLGNIEDYLTVNVCYHLHFTDGTCKPISDLKELLDEIRSFDPVD	126
QY	122	CYSQFMCEQRKNF-AlyPCITRDSYSLKSFLSDLIKALPWLPAPKSVFNMLGYNQE	180
DB	127	GTLRYLAQHERRVQVAREKFIEKSTKPSDFENIDTLCGMOLRLNNMYDDIARFIKD	186
QY	181	KWLAFCFQSKYLGMSPECPALFTMLPYLEHY-GIVHVKGARIIAAMAQVIAENG	239
DB	187	KLRISLTQAYLVGSFPDAPSITYLIGVEHGLSGWYTPKGMNAITQVLKLGEFG	246
QY	240	EHLNSIESLIENGAAKVQLGHGAELRGDEVINADFAHAMTHLVKPGVKYPTEN	299
DB	247	SLRVNAEVQEILLEGRAVGLANGELKADVLSNADFPYTMENTVPASHRGKYTPQ	306
QY	300	LQREYSCSTFWYLGDKIYDLP-HITVAKDYTTNTNFNKLTDDSFYQNAS	358
DB	307	LNSKEVTCGAFMLYLVGNREYNLHNIFTDYKYMDLFTTRQQPFQDPAMTYVRPT	366
QY	359	ASDSLAPAGKALYLVPMPNDSGLDWQAHCQNVREQVLDTLGARLGLSDIRAHICE	418
DB	367	KYDUSVAPPKDIYLVVPVNLSGIDWKETHRYELVIKKL-ERQGVTDLSKHIFE	425
QY	419	KIIPOTWETDEHYVKGATFSLSHKTSOMLYWRPHNRFEELANCYLVGCGTGSLPTI	478
DB	426	RIVTPETFRNFNTYQCAAFGLAPSLFQSgyrPHIKSEVENLYFSGASVHPGGVPV	485
QY	479	YESARISAKLISQK 492	
DB	486	L-----VCGKLVSEQ 495	

RESULT 5  
 Q88SW3 PRELIMINARY; PRT; 498 AA.

ID	Q88SW3	PRELIMINARY; PRT; 498 AA.
AC	Q88SW3	01-JUN-2003 (TrEMBLrel. 24, Created)
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)
DE	Squalene synthase (BC 1.3.-).	
GN	CRTN OR LP 3262.	
OS	Lactobacillus plantarum.	
OC	Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;	
OX	Lactobacillus.	
NCBI_TaxID=1590;		
[1]	SEQUENCE FROM N.A.	
RP	STRAIN-NCIMB 8826 / WCFS1;	
RC	MEDLINE=2480296; PubMed=12566566;	
RX	Kleerebezem W., Boekhorst J., van Kranenburg R., Molenaar D.,	
RA	Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,	
RA	Piers M.W.S.J., Stiekema W., Klein lankhorst R.M., Bron P.A.,	
RA	Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,	
RA	De Vos W.M., Siezen R.J.	
RT	"Complete genome sequence of Lactobacillus plantarum WCFS1."	
RL	Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).	
DR	EMBL; AL953261; CAD5396.1; -.	

DR ProDom: PD139017; Phytin dehydro; 1.  
 DR Hypothetical protein; Complete proc.eome.  
 Q SEQUENCE 488 AA; 55685 MW; F60D539D3CLD49EB CRC64;

Query Match 33.8%; Score 915.5; DB 16; Length 488;  
 Best Local Similarity 37.4%; Pred. No. 1.6e-61;  
 Matches 185; Conservative 95; Mismatches 205; Indels 9; Gaps 2;

1	MANTKHIIIVGAGPGGLCAGMLLSORGFKVSIIDKHAIIGGRNPINMGFTDTGTFL	60
1	MESKKKIATIGAGPGGLATQMLLSQGLQVSYEKNDRIKGGRTALHGKGYSDVGGPSAL	60
61	LMKGVLDMEFELCERSSDYLFPLSPWRLLYDDRDIFVYS-DRENMAEQLORVFDEGT	120
61	TMTHTVLSLFDWCNRLNDVSLPINPIHTLYFKDITFLPYSDVSCTKTIKTYVPPGE	120
121	DGYEQFMCEQRKNFALYPICITRDSYSLKSFLSDLIKALPWLPAPKSVFNMLGYNQE	180
121	DGFDFRENTKOKLYISPLNQPNYSLLDFRPTTIRALPSLATIGRSLMDDLARYNSK	180
181	KWLAFCFQSKYLGMSPECPALFTMLPYLEHYGIYHVKGARIAAMAQVIAENGGE	240
181	YLRIAPSLQWYLGMSPDIPPAAYSIIPESEYFGTFPIGQNKIVZAMQOVVWENKKG	240
241	IHLNSEIESLIENGAAKVQLGHGAELRGDEVINADFAHAMTHLVKPGVKYPTENL	300
241	FYFNESEVTPETNGKETSATLANQSIEADYYFTMLDIFYSLTSNTK---VKE-----	291
301	KOREYSCSTFWYLGDKIYDLPHTHTVAKDYTTNIRNI-PDNKLTDDSFYQNASAS	360
292	KEKEYSSAFMLYLGUKTVLPFHSHQSLIPDPNRFANPTHKKILSKDLAIHTNFSAT	351
361	DDSLAPAGKALYLVPMPNDSGLDWQAHCQNVREQVLDTLGARLGLSDIRAHICEKI	420
352	DNTWAPNHSSVRIWVSAPTTSGLDWKQTPAPRQLILDTVKARLEIPOLEDVIBEEYI	411
421	ITPOTWETDEHYVKGATFSLSHKFSOMLYWRPHNRFEELANCYLVGCGTGSLPTIYE	480
412	ITPTDMEKYHVYGAIFGLHWHOOGLHPAKKSPKFNLFVVYGAGAMSSSLPFIIE	471
481	SARISAKLISQK 494	
472	NAQIATKKFLQEK 485	

RESULT 4  
 Q9ZGE4 PRELIMINARY; PRT; 517 AA.

O9ZGE4	01-MAY-1999 (TrEMBLrel. 10, Created)
O9ZGE4	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
O1-OCT-2003	(TrEMBLrel. 25, Last annotation update)
Diapophytoene dehydrogenase Crtn.	
Crtn.	
Heliobacillus mobilis.	
Bacteria; Firmicutes; Clostridia; Clostridiales; Heliobacteriaceae;	
Heliobacillus.	
NCBI_TaxID=28064;	
[1]	SEQUENCE FROM N.A.
MEDLINE=99061957; PubMed=9843979;	
Xiong J., Inoue K., Bauer C.E.;	
"Tracking molecular evolution of photosynthesis by characterization of	
a major photosynthesis gene cluster from Heliobacillus mobilis.";	
Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).	
PIR; T31463; T31463.	
GO; GO:0016491; P:oxidoreductase activity; IEA.	
GO; GO:0006113; P:electron transport; IEA.	
InterPro; IPR000759; Adnrx_reductase.	
InterPro; IPR002937; Amino_oxidase.	
InterPro; IPR000205; NAD_BS.	
InterPro; IPR008151; Phytin dehydro.	

```
DR GO: 0016491; F: oxidoreductase activity; IEA.
DR GO: 0006118; P: electron transport; IEA.
DR InterPro: IPR000759; Admrx_reductase.
DR PRINTS: PR00419; AdmrdTASE.
DR Oxidoreductase; Complete proteome.
SQ SEQUENCE 498 AA; 56453 MW; C12B1F421DEB31B CRC64;

Query Match 29.8%; Score 808; DB 16; Length 498;
Best Local Similarity 34.4%; Pred. No. 2.8e-53;
Matches 168; Conservative 109; Mismatches 204; Indels 8; Gaps 7;

2Y 5 KHIIVGAGGGLCAGMLLSQSGKVSIFDKAEIGGRNRPINMGFTDTGPTFLMK 64
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 2 KQVSLIAGIGGLTAARVLKGLGNVYIQNAQPGKXNQIVSQFTPDVGFTVWAKD 61
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 65 VLDEMFCERSSDYLEFLPLSPMYLLYDDRDIFVYS-DRENMAEALQVDEGTDGY 123
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 62 IYEEVFRPGVDPHNYLPFEVEQVLMELVFGDSSLDLSDRLPTLIAEINRIAPDDVNGM 121
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 124 EQFWEQERKFNALYPCITRDYSSLSKLSLSDLIKALPWLAPFKSVFNNGQVFNQEK 182
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 122 LNFADLYHRTIAKPNFLERSFNRWDFNLPALYAGLQRTFNNAKYNIAKFNENL 181
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 183 RLAFQSKYLGMSFWPCPALFTMLPYLEHEYGIYHVKGGLNRIAAAMAQVIAENGEIH 242
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 182 RNSLAFQTLYIGISYQGPSLYNIIPMLFLPYGVYFLKGGWYTVNSLVKLFKEQSGTLK 241
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 243 INSIESLIIENGAAKGVKLOHGAELRGDEVINADFAHAMTHLVKGVLLK--YTPENL 300
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 242 LATPVEIITIKDVAVGIRV-NNPIHSDYVLCNADPFTAMTILI-PNEODRQYTDCKI 299
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 301 KOREYSCSTFMYLGLDKIYDLPH-HTIVFAKYDTTIRNIFDKNTLTDDFSFYVONASA 359
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 300 AKMDYSCSLVLYGLDKKYPTTEALHTIHFQADPSKXVADLFQNGKLPEDPSYVYVPSK 359
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 360 SDDSLAPAGKALYVLPMPNDSGLDKQ-AHCQNVREQLVLTGLAELGSLDIRAHICE 418
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 360 MDPSLAPENSEALYVLPVPELSKFNDSPTTLKNYRQLINKRQATPDSIEQIIVVE 419
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 419 KIITPQWETDEHYVKGATPSLSHKFSQMLYWRPHNFEELANCYLVGCGTHPGSGLPTI 478
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 420 KQFTVDKFEQGAINGATGLRPLTKOSNYRPHNKEDYADHLVFCGSSTHFGAGVPIV 479
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 479 YESARISAK 487
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2Y 480 MQSAKLAVE 488
2b : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
Q8YKN8 PRELIMINARY; PRT; 499 AA.
AC Q8YKN8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Zeta-carotene desaturase.
GN ALL7255.
OS Anabaena sp. (strain PCC 7120).
OG Plasmid pCC7120alpha.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa K., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.,
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL; AF003600; BAB78339.1; -.
```

```

R InterPro; IPR000759; Adrnrx_reductase.
R InterPro; IPR008151; Phyt_n_dehydro.
R PRINTS; PR00419; ADXRDTASE.
R PrcDom; PD139017; Phyt_n_dehydro; 1.
Q SEQUENCE 499 AA; 56743 MW; AC1CF06F745B4D34 CRC64;

Query Match.          28.9%; Score 783.5; DB 2; Length 499;
Best local Similarity 35.0%; Pred. No. 2.le-51;
Matches 175; Conservative 103; Mismatches 201; Indels 21; Gaps 10;

Y 4 TKHIIIVGAGPGGLCAGMLLSORGPKVSIIFDKHAEIGGGRNRRPINNMGGFTDTPGTFILMK 63
b 2 SKKVAIVGAGPGGLATAIRLAGYQVEIFEAAERVGGEMRGFEVDSVAEDTGTPIQLP 61

Y 64 GVLDMEFELCERSEYBLEPLPSMYRL-LYDDRRDIFVYSRENRRAELQAVDEGTG 122
b 62 HLYKELPEEAGLNFADYVOLKLEPYTRLKFLWMDGTOLDITSLQSKFKTQLATIRSDLFLA 121

Y 123 YEQFMQERKRNALY-PCITRDYSLKSPLSL-DLIKAL---PMLAPKPSVFNMLGVF 177
b 122 FDRWYSEHIRYHLYGKPYLAGPARSIFGYLRPDDLKFLSPFW---ENLYQHFWEFF 177

Y 178 NQEKMRALAFCSQKLGMSPEWCPALFTMLPYLEHEYGYHYVKGGLNRITAAAMAQVIAEN 237
b 178 QDERLVYDLRYPKYLGMHPTVASSVFSLIPELFSQGVVHFPVGGFRALQAQGLANAADL 237

Y 238 GGEHLNSETESLIIENGAAGKVKIQHGAELRCDEVIINADFAHAMTHLVKPGVLKKYTP 297
b 238 GVKIHLSPVHQIWDQGVQRGLELADASHQGFDTVIINADFAIYVRHLLPITSAGRTID 297

Y 298 ENLKQREYSCSTFMLYGLGDKIY-DLPHTITVPAKDYTTNIRNI----FDNKTLTD-DPS 351
b 298 NKLGMQFSCSTFMLYGINRREYEDLPHQIYL---SDNIRLERPWWDDSGALDSTDP 353

Y 352 FYVONASASDDSLAPAGKSALYVIVMPNRDPSGLDQWQHCVREQVILTLGARIGLSDI 411
b 354 FYVCNPTIIDPSNAPAGHSTFLVLPVTPNTSYAVDWDIKQSYTDFILKRLHL-LGYHNI 412

Y 412 RAHTECKILITPOTWETDEHHVYKGFSLSHKFSQMLYWRPHNRPEELANCYLWGGTHP 471
b 413 EQHIVTOSCTYQASWLDDYRVHLGAVPNLSHNLTLQGFPEPPIRSENIAGLYWIGGAVHP 472

Y 472 GSGLPTIYESARISAKLIISQ 491
b 473 GSGLLTILEASRSAAGFIHQ 492

```

```

RESULT 8
SERNO    PRELIMINARY;      PRT;      511 AA.
D Q8ENMO
C Q8ENMO;
T 01-MAR-2003 (TReMBLrel. 23, Created)
T 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
T 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
E Phytoene dehydrogenase (phytone desaturase) [EC 1.3.-).
N OB2461.
S Oceanobacillus iheyensis.
C Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
X NCBI_X TaxID=182710;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=HTE831 / DSM 14371 / JCM 11309;
X MEDLINE=22220767; PubMed=12235376;
X Takami H., Takaki Y., Uchiyama I.;
T "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
T Ridge and its unexpected adaptive capabilities to extreme
T environments.";
T Nucleic Acids Res. 30:3927-3935 (2002).
L EMBL; AP04601; BAC14417.1;
R GO; GO:0016491; F:oxidoreductase activity; IEA.
R GO; GO:0006118; P:electron transport; IEA.
R InterPro; IPR000759; Adrndx reductase.
R InterPro; IPR002937; Amino Oxidase.

```

```

DR InterPro: IPR000205; NAD_BS.
DR Pfam: PFO1593; Amino oxidase, 1.
DR PRINTS: PR00419; ADXREDTASE.
DR KX: Oxidoreductase; Complete proteome.
DR SQ: SSEQUENCE 511 AA; 58463 MW; 8E70DF9F71D40GAB CRC64;

Query Match 27.3%; Score 739.5; DB 16; Length 511;
Best Local Similarity 32.6%; Pred. No. 5e-48;
Matches 167; Conservative 110; Mismatches 198; Indels 37; Gaps 12;

Qy 5 KHIIVGAPGGLCAGMLLSQGEKVISIFPKHAEIGRRNPINMGCFVDTGPTFLLMKG 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 XKVIVIGAGVAGLASARLQHAGSNVEYKEGMPGCKMHQITKGGTFDGLGFSIVMPE 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 65 VLDEMFELCERRSDYLEFLPLSPMYRLLY----DDRDIFVYSDRENMRALQRFVDSGT 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 LYREVFELTGRNPDYDIPMERIDPMYNVFGNTPEDR-YQISDIAQLIDEIAISDKDA 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 121 DGEVDFEQEQRKRNALYPCITRDYSSLSKSFSLDLKALPLAF-PRKSVFNNL----- 173
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 AGFTYLANKIYQRFQ-----YAKDHFQRFPRH-----PWQFFTPKTLWGLKRLTUG 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 174 -----GOVFNQEKRRALFPCPSQKYLQMSWPECALFTMLPYLEHYGIYHVKGGLNRIA 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 NADNFTKYYIKDRLRQMI SPQTLYIGISPYNGPSLYTIPIMELYGVWFYKGGMYTMA 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 228 AAMAQVIAENGGEIHLANSEIEISLIIENGAAGVKVLOFGAELRGDEVIINADFAHAMTHLV 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 TSNEKLFLLELGGTIHNSPVKEVIIIEKKATGIKLEN-LEINADYVVCNADFPYSMNLI 289
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 288 KQGVLK-KITPENLKQRYSCSTFWLVGLGDKLYDLZH--HTIVPAKYDTYTWIRNIFDNK 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 QDIKAKGKYTDEKIDMKYSCSFIMYLGMKKYKDEVSNNVHNFPSSENKUKETWDDLIPTCK 349
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 345 TLTDQPSFYVQNASDSDSLAPGAKSALYVLVPMNP--NDSGLDW-QAHQCNVREQVLDLT 402
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 350 RL-EDPAFTYIIGSRWDPSLAPEGKOGIYVLVPSVLSLTAQYSGWDETIIAYYREKVFEL 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 403 GARLGLSDIRAHIECEKIITQPTWETDEHVYKCATSLSHKESQMLYWRPHNRFEBLANC 462
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 409 DETDEPKGIKDSVISTETWTLDPDFSEKFNAYNGACPLRPTLSQSHLRPQSKAKNCML 468
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 463 YLVGGGTHPGSLPTIYESARISAKLIISQHR 494
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 469 YFTGSSTHPGAGVPIVLLSAKIAIEELLKDDR 500
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 9
ID Q8KZ28 PRELIMINARY; PRT; 514 AA.
AC Q8KZ28;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Phytoene dehydrogenase.
GN CRTI.
OS uncultured proteobacterium.
OC Bacteria; Proteobacteria; environmental samples.
CX NCBI_TaxID=153809;
RN [1]
RP SEQUENCE FROM M.A.
RX MEDLINE=21822632; PubMed=11832943;
RA Beja O., Suzuki M.T., Heideberg J.P., Nelson W.C., Preston C.M.,
RA Hamada T., Eisen J.A., Fraser C.W., Delong E.F.;
RT "Unsuspected diversity among marine aerobic anoxygenic phototrophs.";
RL Nature 415:630-633(2002).
DR ENBL; AE008921; AAM48646.1; -.
DR GO; GO:0016491; F:oxygenoreductase activity; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR02937; Amino oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR00205; NAD BS.

```



DR InterPro; IPR008151; Phytin dehydro.  
DR Pfam; PF01593; Amino oxidase; 1.  
DR ProDom; PD139017; Phytin dehydro; 1.  
DR PROSITE; PS00982; PHYTOENE DH; 1.  
SQ SEQUENCE 514 AA; 56924 MW; D5177500CAB56773 CRC64;

Query Match 32.9%; Score 729; DB 2; Length 514;  
Best Local Similarity 32.2%; Pred. No. 3.2e-47;  
Matches 160; Conservative 101; Mismatches 210; Indels 26; Gaps 7;

8 IIVGAGPGGLCAGMLLSQSGFKVSIPIFKHAEIGGRNRPINMGFTDTGTFTFLMKGVLD 67  
14 VVIGAGLGLLAAMRLGAKGYATVLDKLRVGGSSVTQDGRFDLGTFTVTMPKVE 73  
68 EMFLCERRSDYLEPLSPMYLLYDDRDIFVYS--DRENMAELQRFVDEGTGVEOF 126  
74 SLMAACGRDHFADVDLRLPLEPFYIRFPDGSYFRAAGDEKMQSEVQRLNPADLPKYRF 133  
127 MEQERKFNALPCIT-----RDYSSLSKSLSLDLKALPWLAPPKSVFNLLQYFNQEK 181  
134 LKDSCKRYIIIGYEGWAEPMRLWETLKVLPFAMLR-----DSIYGLAARRVKDER 187  
182 MRLAFCSQSKYLGMSPWPCPALFTMLPYLEHEYGIYHVKGGLNRIAAQAQVIAENGGEI 241  
188 LRMALSFHPLFIGDPHVTISIALVAHLEKTYGVHYAMGVQIQADAMAAVVRAGQGI 247  
242 HLNSEISLIIENGAAGVKLQHGAEIRGDEVIINADFAHAMTHLVKPGVLYKKTYPENLK 301  
248 HONVADEILIKNGAQAVALTDQRFADPLVSNADAGTYDHLNHSRRRTWTKLA 307  
302 QREYSCSTFMYLIG----LDKIYDLPHTTIVPAKDYTTNIRNIPDNKTLTDDFSPVQNA 357  
308 RKRSMGLFVWYFGTRCTAGRWADVGHITANGPRYKGLLRDIFLKRSLDSDSLVIHRP 367  
358 SASDSDSLAPAGKALYLVPMNN--DSGLDWAQHCNVREOV---LDTLGARLGLSDIR 412  
368 SVTPSPVAPAGDITFYVLSVPVHLGKNKVDQEMPIYAKVAAYEKL-----MEGPE 422  
413 AHIECKIITPQTWETDEHYVKGATFSLSHKFSQMLYWRPHNPFEELANCYLVGGGTHPG 472  
423 ACISTETFTETEDRYLSPHGAGFSIEPILQASAFRPHNVSEARGLYLVGAGTHPG 482  
473 SGLPTIYESARISAKLI 489  
483 AGLPGVSSIAEVLKLV 499

RESULT 10  
Q840T3 PRELIMINARY; PRT; 513 AA.  
AC Q840T3;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Phytoene dehydrogenase.  
GN CRII.  
OS Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Comamonadaceae; Rubrivivax.  
OX NCBI\_TaxID=28068;  
RN SEQUENCE FROM N.A.  
RP STRAIN=SI;  
RC Ouchane S.; Steunou A.-S.; Astier C.;  
RA "Photosynthesis Gene Regulation in Rubrivivax gelatinosus;  
RT Transcription Factor PpsR is Involved in both Negative and Positive  
RT Control";  
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
DR ENBL; AY234385; AAO93135.1; -;  
DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
DR GO; GO:0009058; P:biosynthesis; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR002937; Amino oxidase.

DR InterPro; IPR008150; Bac. phytoene\_dh.  
DR InterPro; IPR000205; NAD\_BS.  
DR InterPro; IPR008151; Phytin dehydro.  
DR Pfam; PF01593; Amino oxidase; 1.  
DR ProDom; PD139017; Phytin dehydro; 1.  
DR PROSITE; PS00982; PHYTOENE DH; 1.  
SQ SEQUENCE 513 AA; 57074 MW; 9C46447BB02F682D CRC64;

Query Match 26.8%; Score 726; DB 2; Length 513;  
Best Local Similarity 32.6%; Pred. No. 5.4e-47;  
Matches 159; Conservative 97; Mismatches 208; Indels 24; Gaps 8;

8 IIVGAGPGGLCAGMLLSQSGFKVSIPIFKHAEIGGRNRPINMGFTDTGTFTFLMKGVLD 67  
24 LVVSGFGGMAAAVRLAAKGYRTVLEKLDAPGGRAYVHRREGHVFDPAGPTIVTPVLPD 83  
68 EMFLCERRSDYLEPLSPMYLLYDDRDIFVYS--DRENMAELQRFVDEGTGVEOF 126  
84 ELWALAGRKFSDDIELKSLDPFYRIREFDDGDHFDYSGDPARMRAEVRRIISPSDAEGFERF 143  
127 MEQERKFNALPCITRDYSSL--KSFLSL-DLIKALP-----WLAFFKSVFNLLQY 176  
144 MREADQ-----CYELGFRTEGDKAPDTVGDLLKAAPLIIRLKG-----RSLHQMVS 192  
177 FNQBRMLAFCSQSKYLGMSPWPCPALFTMLPYLEHEYGIYHVKGGLNRIAAQAQVIAE 236  
193 LKHPKLIAMSLQSLIIGGNPFVSITSMYALVNALBERQWVHWAMGGTGERLIRGLVDVFE 252  
237 N3GEIHLNSETESLIIENGAAGVKLQHGAEIRGDEVIINADFAHAMTHLVKPGVLYKKT 296  
253 MGGTWLKAQVKEVRLEVNGVATGVTLADGERLPADIVVCGNDGTGLYKLVADARWKHT 312  
297 PENLKQREYSCSTFMYLIGLDKIY-DLPHTTIVPAKDYTTNIRNIPDNKTLTDDFSEVQ 355  
313 DARIERGHYSMGLFVWYFGTDRYEDVPHMMVLGPRYRELLDDIFKKKLAGDFSILH 372  
356 NASASDSDSLAPAGKALYLVPMNNDSGLDWAQHCNVREOVLTDLTGARLGLSDIRAH 415  
373 RPTATDPSKAPAGDITFYVLSVPVHLGSGDITWTQAAEPYRQSVQAL-ERTVLPGIQL 431  
416 ECKEIIITPQTWETDEHYVKGATFSLSHKFSQMLYWRPHNPFEELANCYLVGGGTHPGSL 475  
432 RVSFCTTFLDPQHRLLSYKGAGFGLEPLLLQSAFYRPHNRSEDKNLFMVGASTHPGAGV 491  
476 PTIYESAR 483  
492 PGVIMSAR 499

RESULT 11  
Q8JP98 PRELIMINARY; PRT; 511 AA.  
ID Q8JP98  
AC Q8JP98;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Phytoene dehydrogenase.  
GN CRII.  
OS Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Comamonadaceae; Rubrivivax.  
OX NCBI\_TaxID=28068;  
RN SEQUENCE FROM N.A.  
RP STRAIN=IL144;  
RC Nagashima K.V.; Shimada K.; Matsuura K.;  
RA "Phylogenetic analysis of photosynthetic genes of Rhodocyclus  
RT gelatinosus: Possibility of horizontal gene transfer in purple  
RT bacteria";  
RL Photosyn. Res. 36:185-191(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IL144;

X MEDLINE=94132007; PubMed=8300574;  
A Nagashima K.V., Matsuura K., Ohyama S., Shimada K.;  
T "Primary structure and transcription of genes encoding B870 and  
T photosynthetic reaction center apoproteins from Rubrivivax  
T. gelatinosus";  
L J. Biol. Chem. 269:2477-2484 (1994).  
N [3]  
P SEQUENCE FROM N.A.  
C STRAIN=IL144;  
A Igarashi N., Shimada K., Matsuura K., Nagashima K.V.;  
T "Photosynthetic gene cluster in purple bacterium, Rubrivivax  
T. gelatinosus"; (eds.);  
L {in} Garab G. (eds.);  
L Photosynthesis;  
L Mechanisms and effects (Proceedings of the 11th international congress  
L on photosynthesis), pp.4:2889-2892, Kluwer Academic Publishers,  
L Dordrecht (1999).  
N [4]  
P SEQUENCE FROM N.A.  
C STRAIN=IL144;  
X MEDLINE=20031519; PubMed=10563807;  
A Menin L., Yoshida M., Jaquinod M., Nagashima K.V., Matsuura K.,  
A Parot P., Vermeglio A.;  
T "Dark aerobic growth conditions induce the synthesis of a high  
T midpoint potential cytochrome c8 in the photosynthetic bacterium  
T Rubrivivax gelatinosus";  
L Biochemistry 38:15238-15244 (1999).  
R EMBL: AB034704; BA94063.1; -;  
R PIR: T50910; T50910.  
R GO: GO:0016491; P:oxidoreductase activity; IEA.  
R GO: GO:0009058; P:biogenesis; IEA.  
R GO: GO:0006118; P:electron transport; IEA.  
R InterPro: IPR002937; Amino oxidase.  
R InterPro: IPR008150; Bac\_phytoene\_dh.  
R InterPro: IPR002025; NAD\_5S.  
R InterPro: IPR008151; Phyt\_n\_dehydro.  
R Pfam: PF01593; Amino oxidase; 1.  
R ProDom: PD139017; Phyt\_n\_dehydro; 1.  
R PROSITE: PS00982; PHYTOENE DH; 1.  
R SEQUENCE 511 AA; 56618 MW; 9564897196033903 CRC64;  
Query Match 26.7%; Score 725; DB 2; Length 511;  
Best Local Similarity 31.5%; Pred. No. 6.4e-47;  
Matches 158; Conservative 98; Mismatches 208; Indels 24; Gaps 8;

Y 8 IIIVGAGPGGICAGMLLSQKGVKVSIFDKHAEIGGRNRPINMGFTDTGPTLLMKGVLD 67  
b 22 LVVSGFSGGMAAVALAAKGYRVIVLEKLDAPGGRAYVRRHGVDFDAGTTIVTVYFLD 81  
Y 68 EMPCLERSSELYLEPLPSMYRLLYDDRDIFVYS-DRENMRALQRYVDFEG 126  
b 82 ELWALAGKPSDIELKSLDPPFYRIKDFDGDHFDYSGDPARMRAEVRRTSPSDAEGFERF 141  
Y 127 MEQERKRFNLYPCITRDYSSL--KSFSL-DLIKALP-----WLAPPKSVFNNGQY 176  
b 142 MREADQ-----CYELGFTGLDKAFDTVGGDLIKAAPLIVKLRG---RSLHQMVS 190  
Y 177 FNGEKRLAPCFOSKYLGMSPWCPALFTMLPYLEHEGYIVHVGGLNRIAAAMAQVIAE 236  
b 191 LKPKRLIAMSLSLIGGNPFSVTSWYALNVALERQWGVHAWGVTGBLIRGLVDVFEZ 250  
Y 237 NGGEIHLNBSIESLIENGAKGVKLOHGAELRGDEVIINADPAHAMTHLVKPGVLKYYT 296  
b 251 MGSFMRLKAEVRKIEVDNGVATGVTTLADGERIPADIWCNGDTGYLYKLVNDAWRKWT 310  
Y 297 PENLKOREYSCSTFMYLGLDKIY-DLPHHTIVFAKDYTTNIRNFONKTLTDDFSFYVQ 355  
b 311 DALIERGHTSMGLFVWYFGTDREYEDVPHMMVLGPYRELLDDIFRKKCLASDFSILH 370  
Y 356 NASASDDSLAPACKSALYVLVPMNNDSDLDQAHQCNVREQVLDLIGARIGLSIRAH 415  
b 371 RPTATDPSMAPACGDTFYALMPVHGLSGTDWTTQAEYRQSVQEALE-ERTVLPLGQHL 429

QY 416 ECKKITPOTWETDEHYKATFSLSHKPSQMLYRPHNRPEELANCYLVGGTHPGSGL 475  
Db 430 RVSGCTTPLDFQHRLLSYKAGFGLEPLLLQSAFYFRPHNRSEVDKGLFMVGASTHPGAGV 489  
QY 476 PTIVBSAR 483  
Db 490 PGVMSAK 497  
RESULT 12  
Q9KIX2  
ID Q9KIX2 PRELIMINARY; PRT; 506 AA.  
AC Q9KIX2;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Phytoene desaturase.  
GN CRTI.  
OS Bradyrhizobium sp. ORS278  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Bradyrhizobiaceae; Bradyrhizobium.  
OX NCBI\_TaxID=114615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ORS278;  
RX MEDLINE=20309720; PubMed=10951005;  
RA Hannibal L., Lorquin J., Angles d'Ortoli N., Garcia N.,  
RA Chaintreuil C., Masson-Boivin C., Dreyfus B., Giraud E.;  
RT "Isolation and characterization of the canthaxanthin biosynthesis  
RT genes from the photosynthetic bacterium Bradyrhizobium sp. strain  
RT ORS278";  
RL J. Bacteriol. 182:3850-3853 (2000).  
DR EMBL: AF218415; AAF78201.1; -;  
DR GO: GO:0016491; P:oxidoreductase activity; IEA.  
DR GO: GO:0009058; P:biogenesis; IEA.  
DR GO: GO:0006118; P:electron transport; IEA.  
DR InterPro: IPR002937; Amino oxidase.  
DR InterPro: IPR008150; Bac\_phytoene\_dh.  
DR InterPro: IPR008151; Phyt\_n\_dehydro.  
DR Pfam: PF01593; Amino oxidase; 1.  
DR ProDom: PD139017; Phyt\_n\_dehydro; 1.  
DR PROSITE: PS00982; PHYTOENE DH; 1.  
DR SEQUENCE 506 AA; 56223 MW; 435F949E6FDEED3 CRC64;  
Query Match 26.7%; Score 723; DB 2; Length 506;  
Best Local Similarity 31.5%; Pred. No. 9e-47;  
Matches 158; Conservative 107; Mismatches 224; Indels 12; Gaps 8;

QY 1 MANTKHIIIVGAGPGGICAGMLLSQKGVKVSIFDKHAEIGGRNRPINMGFTDTGPTFL 60  
Db 5 MQNKAATVIGSGFGSLSLAIRLQSAGIATTLVKKRDKFGGRAYVTEQDGFDAAGTVI 64  
QY 61 LAKGVLDENFELCERSELYLEPLPSMYRLLYDDRDIFVY-SDRENMRALQRYVDFEG 119  
Db 65 TDPTCQLPALSGRKLNVVELMPVSPFYQLRWEDGATFDYVNDQAELEKQIAAFCPAD 124  
QY 120 TDGYEPMQERKRFNLYPCITRDYSSLKSFSLDLIKALPWL---FPKSVFNNGQY 176  
Db 125 VDGVRFRSPRSEELLEEGY--VKLGHPVDFDPRS--MVRVAPQLVALQSYRSYKVSQY 180  
QY 177 FNGEKRLAPCFOSKYLGMSPWCPALFTMLPYLEHEGYIVHVGGLNRIAAAMAQVIA 235  
Db 181 VSDHILRQAFSFLVGGNPATTSIYALIALERWGVWPRGCTGALINKGLVQLFK 240  
QY 236 ENGGEIHLNBSIESLIENGAKGVKLOHGAELRGDEVIINADPAHAMTHLVKPGVLKYY 295  
Db 241 DLGGEVTLSTSVSRISRIETANGRVSAVVAEDGRERFAADIVASNADVHTYRDLLKDEPLAR 300  
QY 296 TPNLKOREYSCSTFMYLGLDKIY-DLPHHTIVFAKDYTTNIRNFONKTLTDDFSFYV 354  
Db 301 TAQSLMRKFSNLSFYIFGLRREHPELKHILLGRRPRELINIFKGPALPEDFSILY 360  
QY 355 QNASASDDSLAPACKSALYVLVPMNNDSDG-LDQAHQCNVREQVLDLIGARIGLSIDRA 413

Db 361 HAPSVTDSLAPQGGCTYYVLSFVPHLAAPIDWSVEGPRYDRILDVLEARI-LPGKLS 419  
QY 414 HIECEKIITPQWETDEHYKGFATSLSHKFSQMLYWRPHNRFEELANCYLVGCGTHPGS 473  
Db 420 DIATCRITFPDFNFELNAHLGSAFSLPILQTSAYFSAHNDKIKGLYLVGAGTHPGA 479  
QY 474 GLPTIYESARISAKLISOKHR 494  
Db 480 GIPGVVGSATATVLEQDQR 500  
RESULT 13  
Q47845  
ID Q47845 PRELIMINARY; PRT; 492 AA.  
AC Q47845;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Phytoene dehydrogenase.  
GN CRTI.  
OS Pantoea agglomerans.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pantoea.  
OX NCBI\_TaxID=549;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93371414; PubMed=895826;  
RA Liu S.T.;  
RT "Carotenoid-biosynthesis genes as a genetic marker for the purpose of  
gene cloning";  
RL Biochem. Biophys. Res. Commun. 195:259-263(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94236237; PubMed=8180698;  
RA To K.Y., Lai E.M., Lee L.Y., Lin T.P., Hung C.H., Chen C.L.,  
Chang Y.S., Liu S.T.;  
RT "Analysis of the gene cluster encoding carotenoid biosynthesis in  
Erwinia herbicola Ehol3";  
RL Microbiology 140:331-339(1994).  
DR EMBL; M90898; AAA21263.1; -.  
DR PIR; S52586; S52586.  
DR GO; GO:0016491; P:oxidoreductase activity; IEA.  
DR GO; GO:0009058; P:biosynthesis; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR002937; Amino oxidase.  
DR InterPro; IPR008150; Bac phytoene dh.  
DR InterPro; IPR008151; Phyt\_n dehydro.  
DR Pfam; PF01593; Amino oxidase; 1.  
DR ProDom; PD139017; Phyt\_n dehydro; 1.  
DR PROSITE; PS00982; PHYTOENE DH; 1.  
SQ SEQUENCE 492 AA; 55010 MW; 2D65E1A2A32D0635 CRC64;  
Query Match 26.3%; Score 713.5; DB 2; Length 492;  
Best Local Similarity 31.7%; Pred. No. 4.6e-46;  
Matches 158; Conservative 107; Mismatches 208; Indels 25; Gaps 10;  
QY 5 KHIIIVGAGPGCLCAGMLLSQRFKVSIFDKHABIGGRNRPINNGFTFTDGTFLMKG 64  
Db 2 KPTTVIGAGFGLALAIPLQAGIPVLLLEQDKPGGRAYVEDQGTTFDAGTVITDPS 61  
QY 65 VLDEMFICERSESDYLFPLSPMYRLLYDDRIFFVY-SDRENMRALQRFVDEGTDGY 123  
Db 62 AIEELFTLAGQLKDYVELLPVAPFVFLCWESGKVFYNDQAGLEAQIOQFNPRDVEGY 121  
QY 124 EQFVEQERKRNFALYPCITRDYSSLSK----SFLSL-DLIALKALPWA---PPKSVFNGLQ 175  
Db 122 RQFLDYSRAVFEGY-----LKLGVFPLSFRLXLRPAAQLAKLQWRVTVYSKVAS 172  
QY 176 YNQEMRLAFCFOSKYIGMSWPCPALFTMLPYLHEHYGIYHVKGGLNKAAMAQVIA 235  
Db 173 YIEDEHLQAFSFLVGGNPFATSSYITLIHALEREWGVPFRGGTGALVKGMIKLFQ 232

QY 236 ENGGBIHLNSETESLIENGAAKGVKLQHGAEIRGDEVINADFAHAMTHLVK--PGVLK 293  
Db 233 DLGGEVLNAKUSHMETTGDITIEAVHLEDEGRFPTRAVASNADVVHTYRDLLSQHPAAVK 292  
QY 294 KYTPENLKOREYSCSTFMLYGLQKIYD-LPHHTIVFAKDYTTMIRNIFDKNTITDDPSF 352  
Db 293 Q--SKKLQTKRMSNSLFVLYFGLAHHHDQLAHTVCFGRPYRRELIHEIFNEDGLADDFSL 350  
QY 353 YVONASASDDSLAPAGKSAALYVLPMPN-NDSGLDWQAHQCNVREOVLDTLGARLGLSDI 411  
Db 351 YLHAPCVTDSSLAPEGCGSYVLPVPLGTLTANLDWTVEGPRILDRIFEYLEQHY-MPGL 409  
QY 412 RAHIECEKIITPQWETDEHYKGFATSLSHKFSQMLYWRPHNRFEELANCYLVGCGTHP 471  
Db 410 RSLQVTQRMFTPFQDQLNAYQGSFVPELQTSAWFRPHNRDKTINNLYLVGAGTHP 469  
QY 472 GSGLPDIYESARISAKLI 489  
Db 470 GAGIFGVIGSAKATAGLM 487  
RESULT 14  
Q93CI7  
ID Q93CI7 PRELIMINARY; PRT; 504 AA.  
AC Q93CI7;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Phytoene dehydrogenase.  
GN CRTI.  
OS Xanthobacter sp. (strain Py2).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Hyphomicrobiaceae; Xanthobacter.  
OX NCBI\_TaxID=78245;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Larsen R.A., Metcalf W.W.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF408848; AAL02000.1; -.  
DR GO; GO:0016491; P:oxidoreductase activity; IEA.  
DR GO; GO:0009058; P:biosynthesis; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR002937; Amino oxidase.  
DR InterPro; IPR008150; Bac phytoene dh.  
DR InterPro; IPR008151; Phyt\_n dehydro.  
DR Pfam; PF01593; Amino oxidase; 1.  
DR ProDom; PD139017; Phyt\_n dehydro; 1.  
DR PROSITE; PS00982; PHYTOENE DH; 1.  
SQ SEQUENCE 504 AA; 55556 MW; CA848D11A607A828 CRC64;  
Query Match 26.0%; Score 705.5; DB 2; Length 504;  
Best Local Similarity 31.4%; Pred. No. 1.9e-45;  
Matches 154; Conservative 108; Mismatches 211; Indels 17; Gaps 8;  
QY 9 IVGAGPGCLCAGMLLSQRFKVSIFDKHABIGGRNRPINNGFTFTDGTFLMKGVLD 68  
Db 14 VIGSGFGLALAIPLQAGIRTTVEQDKPGGRAYVEQDGTTFDGGTPTVITDPTL 73  
QY 69 MEELCERSESDYLFPLSPMYRLLYDDRIFFVY-SDRENMRALQRFVDEGTDGYEQPM 127  
Db 74 VFAAGRKLSYVELMPVSPFYRLWPDQGFQFVNDQALDAQIATFDKADVEGYRFL 133  
QY 128 EQERKRNFALY-----PCITRDYSSLSKFSLSLKLALPWLAPPKSVFNGLQYFNQEX 181  
Db 134 AYSQAVFEGYKLGAVPFL--EPSSMMK-AAPKLVRLEAW---RSVYAVMSRFIRDDH 186  
QY 182 MRLAFCFOSKYIGMSWPCPALFTMLPYLHEHYGIYHVKGGLNKAAMAQVIAENGGEI 241  
Db 187 LRQAFSFLVGGNPFATSSYITLIHALERKKGVPFRGGTGALVKGMIKLFQ 246  
QY 242 HLNSETESLIENGAAKGVKLQHGAEIRGDEVINADFAHAMTHLVKPGVLKRYTPENLK 301

b 247 RLTPVDIVVEGQRATAVKLKSGETLPLFDLVASNADVVHTYRHLRGAARGRSEGARIA 306  
Y 302 QREYSCSTFMYLGLDKIYD-IPHTTIIPAKDYTTNIRNIFUNKTLTDDPSFYVONASAS 360  
b 307 KTRHNSLSLVTFGARITWDHLQHTVLFPGPRYRGLVDEIFKGPMLPDDFSLYLHAPTPT 366  
Y 361 DQSLAPAGKSALYVLVPMFN-NDSGLDQAHQCNVREOVLDFTIGARLGLSDIRAHIECEK 419  
b 367 DKSLAPEGCTAFYVLPVPHLKGADIDMEVEGPLYRDRILAHLEBRL-LPGLRDSLVTSR 425  
Y 420 ITTPQWETDEHYKGNATSLSHKFSQMLYWRPHRFEELANCYLVGSGTHPGSLPTIY 479  
b 426 ILTPFGFRDLSAHQGSASFVEPLLTQSAWFRPHNRDAKIANLYFAGAGTHPGAGVPGW 485  
Y 480 ESARISAKLI 489  
b 486 GSAKATAGLI 495

RESULT 15  
8GCSI PRELIMINARY; PRT; 492 AA.  
C 08GCSI  
T 01-MAR-2003 (TREMELrel. 23, Created)  
T 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
T 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
E Phytoene desaturase.  
N CRTI.  
S Pantoea stewartii.  
C Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
C Enterobacteriaceae; Pantoea.  
X NCBI\_TaxID=66289;  
N [1]  
P SEQUENCE FROM N.A.  
C STRAIN=ATCC 8200;  
A deSouza M.L., Kollmann S.R., Schroeder W.A.;  
L "Carotenoid Biosynthesis (MO 02/079395 A2).";  
T Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
R EMBL; AY166713; AA85559.1; .  
R GO: 0016491; F:oxidoreductase activity; IEA.  
R GO: 0009058; P:biosynthesis; IEA.  
R GO: 0006118; P:electron transport; IEA.  
R InterPro: IPR002937; Amino oxidase.  
R InterPro: IPR008150; Bac.phytoene\_dh.  
R Pfam: PF01593; Amino oxidase, 1.  
R PROSITE: PS00982; PHYTOENE\_DH; 1.  
Q SEQUENCE 492 AA; 54836 MW; F3DCD224547A5FEA CRC64;

Query Match 25.9%; Score 703.5; DB 2; Length 492;  
Best Local Similarity 31.7%; Pred. No. 2.7e-45;  
Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

Y 5 KHIIIVGAGGGGLCAGMLLSQRFKYSIFDKHAEICGRNRPINMGFTDTGPTFLMKG 64  
b 2 KPTTVIGAGGGGLALAIRQAQIPVLLLEQDKPGGRAYVVOEGFTFDAGPTVITDPS 61  
Y 65 VLDEMPCLERRRDEYLEPLSPMYRLYYDDRDIFVY-SDRENHAEIQRVDEGTDGY 123  
b 62 AIELFALACKQKDVYELLPTVPTFLCWESKVFYNDQAEQAQIQFNPRDVAGY 121  
Y 124 EQFMEQERKFNALYPCITRDYSLK-----SPLSL-DLIKALFWLA---FPKSVFNILQQ 175  
b 122 RAFLDYGRAVFNEGY-----LKLGTVPFLSFKDLRAAPQLAKQAWRSVYSKVAG 172  
Y 176 YFNQEKRLAFQSQKYLGMSPNECPALFTMLPLEHYGIYHVGGLNRIAAQAQVIA 235  
b 173 YIEDEHLRQAFPSHLLVGGNPFATSIYTLIEALERENGVWPPRGTCALVNGAIKLFQ 232  
Y 236 ENGGEIHLNSEIIESLIENGAAGKVLQHGAEALRGDEVIINADFAHAMTHLVK--PGVLK 293  
b 233 DLGGEVVLNARVSHMETVGDKIQAQVLEDGRRETCVASNADVVHTYRDLLSQHPAAAK 292  
Y 294 KYTPENLKQREYSCSTFMYLGLDKIYD-LPHTTIIPAKDYTTNIRNIFUNKTLTDDPSF 352

Search completed: February 29, 2004, 14:50:43  
Job time : 49.0179 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - protein search, using sw model

on on: February 29, 2004, 14:27:18 ; Search time 8.61236 Seconds  
(without alignments)

3089.496 Million cell updates/sec

itle: US-09-941-947a-22

fect score: 2711

quence: 1 WANTKHIIIVAGPGGLGAG.....KRVRFKDIHSAWLKXKA 511

oring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 141681 seqs, 52070155 residues

tal number of hits satisfying chosen parameters: 141681

nimum DB seq length: 0

ximum DB seq length: 200000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	825.5	30.5	529	1 CRTI_MYXXA	Q02861 myxococcus
2	794	29.3	517	1 CRTJ_MYXXA	P54979 myxococcus
3	743	27.4	524	1 CRTJ_RHOCA	P17054 rhodobacter
4	737	27.2	518	1 CRTI_RHOSH	P54980 rhodobacter
5	704.5	26.0	492	1 CRTI_ERWHE	P22871 erwinia her
6	700.5	25.8	492	1 CRTI_PANAN	P21685 pantoea ana
7	686	25.3	583	1 CRTI_PHEBL	P54982 phycomyces
8	681	25.1	595	1 CRTI_NEUCR	P21334 neurospora
9	607	22.4	621	1 CRTI_CERNC	P48537 cercospora
10	605.5	22.3	501	1 CRTI_AGRAU	P54978 agrobacteri
11	508	18.7	507	1 CRTI_STRGR	P54971 streptomyce
12	493	18.2	508	1 CRTI_STRGR	P54971 streptomyce
13	326	12.0	495	1 CRTD_RHOSH	Q01671 rhodobacter
14	325	12.0	494	1 CRTD_RHOCA	P17059 rhodobacter
15	175	6.5	469	1 P45_STRLI	P06108 streptomyce
16	139	5.1	472	1 CRTI_SYNY3	P29273 synechocyst
17	138.5	5.1	393	1 YF88_METVA	Q58883 methanococc
18	134	4.9	500	1 PCL1_ARATH	P57681 arabidopsis
19	129.5	4.8	558	1 ZDS_ARATH	Q38893 arabidopsis
20	128.5	4.7	570	1 CRTI_SOYBN	P28553 glycine max
21	126.5	4.7	516	1 OXLA_AGRKH	P31392 sgkistrodon
22	126.5	4.7	588	1 ZDS_CAPAN	Q95m33 capsicum an
23	125.5	4.6	587	1 ZDS_TARER	Q91v46 tagetes ere
24	123	4.5	516	1 OXLA_CROAD	Q93364 crocalus ad
25	120.5	4.4	471	1 GLTD_ECOLI	P09832 escherichia
26	119.5	4.4	588	1 ZDS_LYCES	Q9se20 lycopersico
27	118.5	4.4	583	1 CRTI_LYCES	P28554 lycopersico
28	117.5	4.3	445	1 GDIC_RAT	P50399 rattus norv
29	117	4.3	749	1 TR2M_AGRRH	Q09103 agrobacteri
30	115.5	4.3	447	1 GDIA_HUMAN	P31150 homo sapien
31	115.5	4.3	447	1 GDIA_PANTR	P60028 pan troglod
32	115.5	4.3	582	1 CRTI_CAPAN	P80093 capsicum an
33	115	4.2	566	1 CRTI_ARATH	Q07356 arabidopsis

34	114.5	4.2	445	1 GDIC_MOUSE	Q61598 mus musculu
35	114.5	4.2	447	1 GDIA_CANFA	O97555 canis famil
36	114.5	4.2	447	1 GDIA_MOUSE	P50396 mus musculu
37	114.5	4.2	447	1 GDIA_RAT	P50398 rattus norv
38	114.5	4.2	613	1 YC68_CAMJE	Q9pn30 campylobact
39	112.5	4.1	445	1 GDIB_HUMAN	P50395 homo sapien
40	112.5	4.1	447	1 GDIA_BOVIN	P21856 bos taurus
41	112.5	4.1	639	1 YGFT_ECOLI	Q46820 escherichia
42	111	4.1	471	1 PPOX_MYXXA	P56601 myxococcus
43	111	4.1	508	1 FMS1_YEAST	P50264 saccharomyc
44	109.5	4.0	566	1 CRTI_ORYSA	Q9ztng oryza sativ
45	109.5	4.0	574	1 ZDS_ORYPS	O49901 narcissus p

#### ALIGNMENTS

##### RESULT 1

ID	CRTI_MYXXA	STANDARD;	PRT;	529 AA.
AC	Q02861;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).			
GN	CARC.			
OS	Myxococcus xanthus.			
OC	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;			
OC	Cystobacterineae; Myxococcaceae; Myxococcus.			
OX	NCBI_TaxID=34;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MR403;			
RX	MEDLINE=93223667; PubMed=8467787;			
RA	Fontes M., Ruiz-Vazquez R.M., Murillo F.J.;			
RT	"Growth phase dependence of the activation of a bacterial gene for			
RT	carotenoid synthesis by blue light."			
RL	EMBO J. 12:1265-1275(1993).			
CC	!- FUNCTION: This enzyme converts phytoene into lycopene via the			
CC	intermediaries of phytofluene, zeta-carotene and neurosporene by			
CC	the introduction of four double bonds (By similarity).			
CC	!- COFACTOR: FAD (Probable).			
CC	!- PATHWAY: Carotenoid biosynthesis.			
CC	!- INDUCTION: By blue light.			
CC	!- SIMILARITY: Belongs to the phytoene dehydrogenase family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M94727; AAA25390.1; ..			
DR	PIR; S35306; S35306.			
DR	InterPro; IPR000759; Adrmdx reductase.			
DR	InterPro; IPR002937; Amino oxidase.			
DR	InterPro; IPR008150; Bac phytoene dh.			
DR	InterPro; IPR000205; NAD BS.			
DR	InterPro; IPR008151; Phytn dehydro.			
DR	Pfam; PF01593; Amino oxidase; 1.			
DR	PRINTS; PR00419; ADXDRPASE.			
DR	ProDom; PD139017; Phytin dehydro; 1.			
DR	PROSITE; PS00982; PHYTOENE DH; 1.			
KW	Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.			
FT	NP BIND 12 45 FAD (ADP PART) (POTENTIAL).			
SQ	SEQUENCE 529 AA; 58420 MW; 5353688DFD0D24BC CRC64;			

Query Match 30.5%; Score 825.5; DB 1; Length 529;

Best Local Similarity 35.1%; Pred. No. 2.1e-54;

Matches 179; Conservative 106; Mismatches 192; Indels 33; Gaps 10;



Armstrong G.A., Alberti M., Leach P., Hearst J.E.;  
"Nucleotide sequence, organization, and nature of the protein  
products of the carotenoid biosynthesis gene cluster of Rhodobacter  
capsulatus.";  
Mol. Gen. Genet. 216:254-268(1989).  
[2]  
SEQUENCE FROM N.A.  
MEDLINE=89327279; PubMed=2546948;  
Bartley G.E., Scolnik P.A.;  
"Carotenoid biosynthesis in photosynthetic bacteria. Genetic  
characterization of the Rhodobacter capsulatus CrtI protein.";  
J. Biol. Chem. 264:13109-13113(1989).  
[3]  
ERRATUM.  
Bartley G.E., Scolnik P.A.;  
J. Biol. Chem. 264:18260-18260(1989).  
[4]  
SIMILARITY TO CAROTENOID DESATURASES.  
MEDLINE=90368827; PubMed=2144293;  
Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;  
"Carotenoid desaturases from Rhodobacter capsulatus and Neurospora  
crassa are structurally and functionally conserved and contain  
domains homologous to flavoprotein disulfide oxidoreductases.";  
J. Biol. Chem. 265:16020-16024(1990).  
[5]  
FUNCTION: This enzyme converts phytoene into zeta-carotene via the  
intermediary of phytofluene by the symmetrical introduction of two  
double bonds at the C-11 and C-11' positions of phytoene.  
[6]  
COFACTOR: FAD (Probable).  
[7]  
PATHWAY: Carotenoid and chlorophyll biosynthesis.  
[8]  
SIMILARITY: Belongs to the phytoene dehydrogenase family.  
[9]  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
[10]  
EMBL; J04969; AAA50313.1; -;  
EMBL; X52291; CAA36533.1; -;  
EMBL; Z11165; CAA77540.1; -;  
PIR; A32617; A32617.  
InterPro; IPR000759; Amdrxn reductase.  
InterPro; IPR002937; Amino oxidase.  
InterPro; IPR008150; Bac\_phytoene\_dh.  
InterPro; IPR000205; NAD\_B8.  
InterPro; IPR008151; Phytin dehydro.  
Pfam; PF01593; Amino oxidase; 1.  
PRINTS; PR00419; ADXRD7ASE.  
ProDom; PD139017; Phytin dehydro; 1.  
PROSITE; PS00982; PHYTOENE DH; 1.  
Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;  
Oxidoreductase; FAD; Flavoprotein; NAD.  
NP BIND 12 45 FAD (ADP PART) (POTENTIAL).  
SEQUENCE 524 AA; 57978 MW; 6425A7ESA06AA6B9 CRC64;

QY 183 RLAFQFQSKYLGMSPWECFALFTMLPYLEHYGIYHVKGGLNRIAAAMQVIAENGSGIH 242  
Db 187 RPAISFHLPLFGDDPFHVTSMVILVSQLEKKFGVHYAIGGVQVQAIADAMAKVITDQGGEMR 246  
QY 243 LNSEIESLII-ENGBAKGVKLOHGAELGDEVINADFAHAMTHLYKPGVLKKTYPENLK 301  
Db 247 LNTVEDEILVSRDGGKATGIRLMDGTLEPAQVNVNADAGHTYTKLLNRDRWRWTDEKLD 306  
QY 302 QREYSCSTFMLYLGLD---KIY-DLPHTITVFAKDYTTNIRNIFDNKTLTDDFSFYQNA 357  
Db 307 KRWEMGLFWYFGTKAKWKQDVGHVTVVGGRYKEHVQDIFIKGELAEDMSLYVHRP 366  
QY 358 SASDLSLAPAGKSAIYLVPMN--NDGSLDQWQAHQNVREQLDITLGLARGLSDIRAH 415  
Db 367 SVTDPTAAPKGDITFYLSPVNLGDFNGVDSVEAEKYKAKVLKVIIBERL-LPGVAKKI 425  
QY 416 ECEKIITPOTWETDEHYVKGATFSLSHKFSQMLYWRPHNRPEELANCYLVGCGTHPGSG 475  
Db 426 TEVVTFTETFRDYLSPGAGFSLPEILLOSANFRPHNASEEVDGLYLVGAGTHPGAGV 485  
QY 476 PTIYESARISAKLI 489  
Db 486 PSVIGSGELVAQMI 499

RESULT 4

CRTI\_RHOSH  
ID CRTI\_RHOSH STANDARD; PRT; 518 AA.  
AC P54980; QSRFDO;  
DT 01-OCT-1996 (Rel. 34; Created)  
DT 01-OCT-1996 (Rel. 34; Last sequence update)  
DE 28-FEB-2003 (Rel. 41; Last annotation update)  
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).  
GN CRTI.  
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
OC Rhodobacteraceae; Rhodobacter.  
OX NCBI\_TaxID=1063;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;  
RX MEDLINE=94292403; PubMed=8021167;  
RA Lang H.P., Cogdell R.J., Gardiner A.T., Hunter C.N.;  
RT "Early steps in carotenoid biosynthesis: sequences and  
transcriptional analysis of the crtI and crtB genes of Rhodobacter  
sphaeroides and overexpression and reactivation of crtI in  
Escherichia coli and R. sphaeroides.";  
RL J. Bacteriol. 176:3859-3869(1994).  
RN [2]  
SEQUENCE FROM N.A.  
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;  
RX MEDLINE=95238278; PubMed=7721699;  
RA Lang H.P., Cogdell R.J., Takaichi S., Hunter C.N.;  
RT "Complete DNA sequence, specific Tn5 insertion map, and gene  
assignment of the carotenoid biosynthesis pathway of Rhodobacter  
sphaeroides.";  
RL J. Bacteriol. 177:2064-2073(1995).  
RN [3]  
SEQUENCE FROM N.A.  
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;  
RX MEDLINE=20115311; PubMed=10648776;  
RA Choudhary M., Kaplan S.;  
RT "DNA sequence analysis of the photosynthesis region of Rhodobacter  
sphaeroides 2.4.1.";  
RL Nucleic Acids Res. 28:862-867(2000).  
CC -!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the  
intermediary of phytofluene by the symmetrical introduction of two  
double bonds at the C-11 and C-11' positions of phytoene.  
CC -!- COFACTOR: FAD (Probable).  
CC -!- PATHWAY: Carotenoid biosynthesis.  
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

Query Match 27.4%; Score 743; DB 1; Length 524;  
Best Local Similarity 32.8%; Pred. No. 3.2e-48;  
Matches 162; Conservative 103; Mismatches 211; Indels 18; Gaps 9;  
8 IIVGACPGGLCAGMLLSQGRFKVIFDKHAEIGGNRPINNNGFTPTGTFTLLKGVLD 67  
12 VVIGAGLGLAAANRLGAKYKVTVDRLDRPGRGSSITKGGHRTGTIVTPDRLR 71  
68 EMFELCERSESDYLEFLPLSMYRLLYDDRIDP-VYSDRENRAELORVDEGTDGVEQF 126  
72 ELWADCGDFDFOVSLVPWEPFYITDPPDGEKTYTAGDAAKVAARVARISPGDVEGPRHF 131  
127 MEQEKREKNALYPCITRUYSSLSKFLSL-DLIKALP---WLAFPKSVFNLLGQVFNQERM 182  
132 MWDAKARVEFGYENLGR-----KPMKSLWDLIKVLPFTFGWLADRDSYVGHAKWVKDHL 186

```

C between the Swiss Institute of Bioinformatics and the EMBL outstation
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (See http://www.eib-sib.ch/announce/
C or send an email to license@sib-sib.ch).
C -----
C EMBL; S71770; AAB31138.1; --
C EMBL; AJ010302; CAB38739.1; --
C EMBL; AF195122; AAF24289.1; --
C PIR; S49620; S49620.
C PIR; T50745; T50745.
C InterPro; IPR002937; Amino oxidase.
C InterPro; IPR008150; Bac phycoene_dh.
C InterPro; IPR002025; NAD_BS.
C InterPro; IPR008151; Phycn_dehydro.
C Pfam; PF01593; Amino oxidase; 1.
C ProDom; PD139017; Phyt dehydro; 1.
C PROSITE; PS00982; PHYTOENE DH; 1.
C Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
C Oxidoreductase; FAD; Flavoprotein; NAD.
C NP_BIND 14 47 FAD (ADP PART) (POTENTIAL).
C CONFLICT 73 73 R -> C (IN REF. 3).
C CONFLICT 174 174 T -> S (IN REF. 3).
C CONFLICT 292 292 L -> F (IN REF. 3).
C CONFLICT 395 395 Q -> P (IN REF. 3).
C SEQUENCE 518 AA; 57244 MW; 423B3515639BP8P1 CRC64;
Query Match 27.2%; Score 737; DB 1; Length 518;
Best local Similarity 33.5%; Pred. No. 8.9e-48;
Matches 170; Conservative 97; Mismatches 207; Indels 34; Gaps 10;
Y 8 II VAGSGGGLACGMLLSQSGFKVSIIPDKHAEITGGRRRPNINWNGTFTDTGPTFLLMKGVLID 67
D 14 LVIGSGLGLAAAWLGLGKGRVTWIDKLDVPGRGSSITQEGHREFDLGPTIVTVPOSIR 73
Y 68 EMFLCERRSEDYLEFLPLSPMYRLLYDDRDIF - VYSDRNMRAELQRVPDGDGTGTEQF 126
D 74 DLKTCGRDFDADVELKPIDPFYEYEWPDGSHFTYRQSTEAHKAFAVLRLSPGDVAGYEKF 133
Y 127 MEQKREKFNALYPCITRDYSSYLKSLFLSL-DLIALKP- --WLAFPKSVFNLLGQYFNQEKM 182
D 134 LKQSEKRYWFGYEDLGR-----RSMKMLWDLIKVLPTFGMRADRTVYQHAALRVKDERL 188
Y 183 RLAFCFOSKYLGMSPWECPALFTMLPYLEBEHYGIYHVKGGLNRJAAAMAQVIAENGGEIH 242
D 189 RMALSPHPLFIGGDPFNVTSMYILVSOLEKEFGVHVAITGGVAAATAAMAKVIEQGGSGFR 248
Y 243 LNSHIESLIITENAAKGVKLQHCAGELRGDEVIINADFAHMTHLVKGVLKKTYPENLKQ 302
D 249 MNTBVDSEILVEKGTATGVRLASGEVLRAGLVVSNADAGHTYMRLLRNPRRRTWDADAVKS 308
Y 303 REYSCSTFMFLYGLDKTY ----DLPHHTIVFAKDYYTIRNIFDNKTLTDDFSFYVQNAS 358
D 309 RMWSMGLFVWYFGTKGKWMPDVGHHTIVNAPRYKGLVEDIFLKGKLAKMSLYIHRPS 368
Y 359 ASDSLAPAKGSALYLVNPMFN--NDSGLDWAHQHQNREQVLDL-----CARGLLS 409
D 369 ITDPTVAPEGDDTFYALSPVPHLQKQPDQWQAVAPYRESYLEVLEQSMFGIGERIGPS 428
Y 410 DIRAHIECEKIIIPQWTEWTEHYVK-GATFSLSHKPSQMLYVRPHNRPEELANCYLVGGG 468
D 429 -----LVFTPEIFR-DRLYSLPWGAGFSETEPRILASAWFRPHNISEEVANLFIWGAG 478
Y 469 THPGSGLPETTESARISAKLISQKHVR 496
D 479 THPGAGVPGVIGSAEYMAKLADAPRAR 506
RESULT 5
RTTI ERNHE
ID CRTI ERWHE STANDARD; PRT; 492 AA.
AC P22871;

```

```

01-AUG-1991 (Rel. 19, Created)
01-AUG-1991 (Rel. 19, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).
CRTO.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=EHO10;
RM MEDLINE=9108634; PubMed=2263648;
RA Armstrong G.A., Alberti M., Hearst J.E.;
RT "Conserved enzymes mediate the early reactions of carotenoid
RT biosynthesis in nonphotosynthetic and photosynthetic prokaryotes.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:9975-9979(1990).
CC -!- FUNCTION: This enzyme converts phytoene into lycopene via the
CC intermediaries of phytofluene, zeta-carotene and neurosporene by
CC the introduction of four double bonds.
CC -!- COPACTOR: FAD (Probable).
CC -!- PATHWAY: Carotenoid biosynthesis.
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M38423; AAA24820.1; -
DR EMBL; M67280; AAA64981.1; -
DR PIR; A39273; A33120.
DR InterPro; IPR002937; Amnio oxidase.
DR InterPro; IPR008150; Bac phytoene dh.
DR InterPro; IPR008151; Phyt_n_dehydro.
DR Pfam; PF1593; Amnio.Oxidase; 1.
DR ProDom; PD39017; Phyt_n_dehydro; 1.
DR PROSITE; PS00982; PHYTOENE_DH; 1.
DR Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.
FT NP BIND 5 38 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 492 AA; 54503 MW; 8EDC5DB1562083F2 CRC64;
Query Match 26.0%; Score 704.5; DB 1; Length 492;
Best Local Similarity 31.7%; Pred. No. 2.3e+45;
Matches 158; Conservative 102; Mismatches 213; Indels 25; Gaps 9;
QY 5 KHTIIVGAGPGGLCAGMLISQRGEKVSIPDKFAEIGGRNRPINMGFTDTGPTFLLMKG 64
DB : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 2 KKTTVIGAGFGGLALAIRLQAAGIPTVLLEQRDKGGRAYVWHDAQGTTFDAGPTVTDP 61
QY 65 VLDEMELCEKRSSELYELPLSPMYRLLDYDRDFIVYSDRENMAELQRVEDECTGY 123
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 ALAELFTLAGRMEDYVRLLPKFPFYRLCWESGKTLDYANDSAELEAQITQNPRDVEGY 121
QY 124 EQFMEOERKFNALYPCITRDYSLLKS--FLSL-DLIKALP-----WLAFPKSVFNLL 173
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 122 RRLAYSQAVFO-----EGYLRLGSVFLLSPROMLRAGPOLKLQAM----QSYYQSV 170
QY 174 GOYFNOKRRRLAPCFOSKYLGMSWPECALFTMLPYLEHYGIYHVKGLNRFAAMAQV 233
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 171 SRFIEHILRQAFSFEHLVGGNPFTTSIYTLIALEREWGVPPEGGTGALVNGMKVL 230
QY 234 IAENGCEIHLNGEIESLLIENGAAGVKQLHQGAELRGDEVINADPAHAMTHLVKPGVLK 293
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 231 FTDLGGEIELNARVELVADVADRVSQVRLADGREIFDDTAVASNAVYVNTYKKLGHHPVG 290
QY 294 KYTPENLKQREYSCSFMYLGLDKIY-DLPHTTIYFAKDYTNNIERNIDNKLTLDFFSF 352
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 291 QKKAALAEKRKMNSLFVIYGLGNOPHSOLAHTICGFERYRELIDEIFTGSALADDPSL 350
QY 234 IAENGCEIHLNGEIESLLIENGAAGVKQLHQGAELRGDEVINADPAHAMTHLVKPGVLK 293
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 231 FTDLGGEIELNARVELVADVADRVSQVRLADGREIFDDTAVASNAVYVNTYKKLGHHPVG 290
QY 294 KYTPENLKQREYSCSFMYLGLDKIY-DLPHTTIYFAKDYTNNIERNIDNKLTLDFFSF 352
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 291 QKKAALAEKRKMNSLFVIYGLGNOPHSOLAHTICGFERYRELIDEIFTGSALADDPSL 350

```

RESULT 5  
TRTI\_ERWH  
ID-CRTI  
AC P228







01-FEB-1996 (Rel. 33, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).  
PDB1.  
Cercospora nicotianae.  
Eukaryota; Fungi; Ascomycota; Pezizomycotina;  
Dothideomycetes et Chaetothyrionycetes incertae sedis;  
Mycosphaerellaceae; Mitosporic Mycosphaerellaceae; Cercospora.  
NCBI\_TaxID=29003;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=ATCC 18366;  
MEDLINE=94368091; PubMed=8085820;  
Zehrfahrt M., Daub M.E.;  
"Isolation, sequence, and characterization of the Cercospora  
nicotianae phytoene dehydrogenase gene";  
Appl. Environ. Microbiol. 60:2766-2771(1994).  
-!- FUNCTION: This enzyme converts phytoene into zeta-carotene via the  
intermediary of phytofluene by the symmetrical introduction of two  
double bonds at the C-11 and C-11' positions of phytoene.  
-!- COFACTOR: FAD (Probable).  
-!- PATHWAY: Carotenoid biosynthesis.  
-!- SIMILARITY: Belongs to the phytoene dehydrogenase family.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; U03903; AAB86388.1; --  
PIR; T48646; T48646.  
InterPro; IPR008150; Bac\_phytoene\_dh.  
InterPro; IPR008151; Phyt\_dhydro.  
ProDom; PD139017; Phyt\_dhydro; 2.  
PROSITE; PS00982; PHYTOENE DH; 1.  
Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD;  
Transmembrane.  
NP\_BIND 11 44 FAD (ADP PART) (POTENTIAL).  
TRANSMEM 536 556 POTENTIAL.  
SEQUENCE 621 AA; 59529 MW; 6189EA784963C8A8 CRC64;  
Query Match 22.4%; Score 607; DB 1; Length 621;  
Best Local Similarity 27.2%; Pred. No. 6.5e-38;  
Matches 151; Conservative 114; Mismatches 211; Indels 80; Gaps 11;  
8 IIIVGAPGGLCAGMLLSQSGKVSIFDKHAEIGGRNPINNGTFTGTFTFLMKGVLD 67  
11 IIIVGSGVGVSTAAKRLARAGHVTVLEKNFTGRCSLIHHEGYRFDQGSLLLLPLGFH 70  
68 EMP-ELCERRSEDYLEFLPLSPMYRLAYDRDIF-VYSDRENMAELQRFVDEGTGQEY 125  
71 RTFABLTSLQEGVKLLKCPNMYIHFSDEKFTLSDSLVMKTEVKW--EGKGYTR 128  
126 FMEQKRFN---ALYICITRDYSSLSFLSLDLIKALPWLAPPKSVFNQLQYFNQEK 181  
129 YLEFLKSHGHYELSVREVLNREFTGLTAMLRPFERHLLQLHPFESINTRAGKYFWTER 188  
182 MRLAPCFOSKYLGNMPCAPFLTWLPYLEHYGYHVKGGLNRIAAQAQVLAENGEL 241  
189 LERVFTFOSMYNGMSPDPAGTYSILQYTELAEIGIWPVGSHRVEALVKIGERGVDF 248  
242 HUNSEIESLII--ENGAAGKVLQHGAELEGDVVIINADFAHAMTHLVKPGVLKXYPEN 299  
249 RMETAVKILLSDEGVAKGVELEGRLEADVVVNNSDLVAYAEKLLP---IKTPYAES 305  
300 LKQREYSCSTFWLYGLDK-IVDLPHHTIVFAKDYTTNIRNIDNKLTLDPSFYQNAS 358  
306 LKRGCGSCSSISFTWALDQRFLEAHNIFLADEYRESFOSIFKXHLIPDEPSFYNVPS 365  
359 ASDDSLAPAGKSAIYVLYPM-----PNNDSGL----- 385

Db 366 RVDSTAAPRGKDSVVVLPVGVHLEEDRHASQAHLASRNHGHISSASPPDQGLTPTEK 425  
Qy 386 -DQAHQCNVREQVLDLTGARGLSDIRAHIECEKIITPQWETDEHYVKGATFSLSHKF 444  
Db 426 QDWPAMISLARKTILSTIQSRNV-DLTLIHSTNSPLSMKQTFNDRGAILGLSHSF 484  
Qy 445 SQMLYWRP-----HNRFEELANCYLVGGGTHFGS 473  
Db 485 FNVLCFRPTTRAKPGAFDAQLLKFGVLGRAEVIIDAFGRGKDIKGLVYVYVYVYVYV 544  
Qy 474 GLPTIYESARISAKLI 489  
Db 545 GVPICLAGGALVARGI 560  
RESULT 10  
CRTI AGRAU  
ID CRTI AGRAU STANDARD; PRT; 501 AA.  
AC P54978;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Phytoene dehydrogenase (EC 1.14.99.-) (Phytoene desaturase).  
GN CRTI.  
OS Agrobacterium aurantiacum.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
OX NCBI\_TaxID=44155;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96062243; PubMed=7592436;  
RA Misawa N., Satomi Y., Kondo K., Yokoyama A., Kajiwara S., Saito T.,  
RA Ohtani T., Miki W.;  
RT "Structure and functional analysis of a marine bacterial carotenoid  
biosynthesis gene cluster and astaxanthin biosynthetic pathway  
proposed at the gene level";  
RL J. Bacteriol. 177:6575-6584(1995).  
CC -!- FUNCTION: This enzyme converts phytoene into lycopene via the  
intermediaries of phytofluene, zeta-carotene and neurosporene by  
the introduction of four double bonds (By similarity).  
CC -!- COFACTOR: FAD (Probable).  
CC -!- PATHWAY: Carotenoid biosynthesis. Involved in astaxanthin  
biosynthetic pathway.  
CC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D58420; BAA09594.1; --  
DR InterPro; IPR002937; Amino oxidase.  
DR InterPro; IPR008150; Bac\_phytoene\_dh.  
DR InterPro; IPR008151; Phyt\_dhydro.  
DR Pfam; PF01593; Amino oxidase; 1.  
DR ProDom; PD139017; Phyt\_dhydro; 1.  
DR PROSITE; PS00982; PHYTOENE DH; 1.  
KW Carotenoid biosynthesis; Oxidoreductase; FAD; Flavoprotein; NAD.  
NP\_BIND 12 45 FAD (ADP PART) (POTENTIAL).  
SQ SEQUENCE 501 AA; 54806 MW; 5F251AF1D679358 CRC64;  
Query Match 22.3%; Score 605.5; DB 1; Length 501;  
Best Local Similarity 28.6%; Pred. No. 6.3e-38;  
Matches 141; Conservative 110; Mismatches 227; Indels 15; Gaps 8;  
Qy 5 KHIIVGAPGGLCAGMLLSQSGKVSIFDKHAEIGGRNPINNGTFTGTFTFLMKG 64  
Db 9 KTAIVIGAGFGGLALIRLQSGIATTLVEARDKPGGRAYVYVYVYVYVYVYVYVYV 68





```

QY 164 APPKSVFNNLGOYFNQEKORLAPCFQSKYLQMSWPCALFTMLPYLEHEYGIYHVKGGL 223
Db 182 -----FFDRRLRLQFRYATYVGTGYPAGVGLALIT-WAAEARGVWAEIGGM 228
QY 224 NRIAAAOVIAENGGEIHLNSETESIENGAAGKGVKLOHGAELRGDEVLINADFAHAM 283
Db 229 HRLALALARLADDOGVRLRYCAPVAGILRQGRRTGVQLADGRTLPADHIVFNGDPAALL 288
QY 284 THLVKPGVLKYYTENIKQREYSCSTFWMLYGLDKIYDLPHHTIVFAXDYTTNIRIFDN 343
Db 289 AGCLGDGQDQAVPDRIPRSLSAWWS-YAARASGPPVLVHNVFFADDPREFGPIAAG 347
QY 344 KTLTDDREFFYQNASADSLAPACKSALYVLPVPMNDSGLDQAHQCNVQVLDTLG 403
Db 348 Q-MEDAILYICADRSQGLPDGPPELIIINGPDPORPAKPSDPAC-----RSRTFRL- 402
QY 404 ARLGLSDIRAHIECKIITPOTWETDEHYKAGTAFSLKFSQMLYWRPNRFEELANCY 463
Db 403 RQFGLTFDPVGE-TSLTAPSGFASLPASQGSYGLSPHGALASLKRPLAR-TALPGLW 460
QY 464 LVGSGTHGSGLPITYESARISAKLI 489
Db 461 LAGGAHPGAGVPMALSGRHAERAI 486

RESULT 14
RTD_RHOCA
ID_CRTD_RHOCA STANDARD; PRT; 494 AA.
YC P17059;
YT 01-AUG-1990 (Rel. 15, Created)
YT 01-AUG-1990 (Rel. 15, Last sequence update)
YT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Methoxynuroporene dehydrogenase [EC 1.14.99.-].
ZS CRTD.
ZS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
XC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
XC Rhodobacteraceae; Rhodobacter.
XX NCBI_TaxID=1061;
ZP SEQUENCE FROM N.A.
ZP STRAIN=SB1003 / St Louis, and BEC404;
ZP MEDLINE=89313663; PubMed=2747617;
ZP Armstrong G.A., Alberti M., Leach F., Hearst J.E.;
RT "Nucleotide sequence, organization, and nature of the protein
RT products of the carotenoid biosynthesis gene cluster of Rhodobacter
RT capsulatus."
ZL Mol. Gen. Genet. 216:254-268(1989).
ZP [2]
ZP SIMILARITY TO CAROTENOID DESATURASES.
ZP MEDLINE=90368827; PubMed=214293;
ZP Bartley G.E., Schmidhauser T.J., Yanofsky C., Scolnik P.A.;
RT "Carotenoid desaturases from Rhodobacter capsulatus and Neurospora
RT crassa are structurally and functionally conserved and contain
RT domains homologous to flavoprotein disulfide oxidoreductases."
ZL J. Biol. Chem. 265:16020-16024(1990).
ZC -!- FUNCTION: CONVERTS HYDROXYNEUROSPORENE TO DEMETHYLSPHEROIDENE OR
ZC METHOXYNEUROSPORENE TO SPHEROIDENE.
ZC -!- COFACTOR: FAD (Probable).
ZC -!- PATHWAY: Carotenoid and chlorophyll biosynthesis.
ZC -!- SIMILARITY: Belongs to the phytoene dehydrogenase family.
ZC This SWISS-PROT entry is copyright. It is produced through a collaboration
ZC between the Swiss Institute of Bioinformatics and the EMBL outstation -
ZC the European Bioinformatics Institute. There are no restrictions on its
ZC use by non-profit institutions as long as its content is in no way
ZC modified and this statement is not removed. Usage by and for commercial
ZC entities requires a license agreement (See http://www.isb-sib.ch/announce/
ZC or send an email to license@isb-sib.ch).
ZC -----
ZC EMBL: X52291; CAA36537.1; --
ZC EMBL: Z11165; CAA77544.1; --
ZC PIR: S04406; S04406.
ZC InterPro: IPR002937; Amino oxidase.

```

```

DR InterPro: IPR008150; Bac_phytoene dh.
DR InterPro: IPR008151; Phytan dehydro.
DR Pfam: PF01593; Amino oxidase; 1.
DR PROSITE: PD139017; Phytan dehydro; 1.
DR PROSITE: PS00982; PHYTOENE DH; 1.
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis;
KW Oxidoreductase; FAD; Flavoprotein; NAD.
ET NP_BIND 8
SQ SEQUENCE 494 AA; 52312 MW; D1180A023FFB5A9 CRC64;

Query Match
Best Local Similarity 12.0%; Score 325; DB 1; Length 494;
Matches 127; Conservative 23.9%; Pred. No. 7e-17;
Matches 127; Indels 88; Gaps 21;

QY 1 MANTKHIIIVAGAGGGLCAGMLLSQRFKYSIFDKHAEIGGRNRPINMGFTPTGTGTFPL 60
Db 1 KRSETDVVVGARMGGLAAAGIAAGAAAGLVTVVEAGDAPGGKARAVTPOGPADTGTPTVL 60
QY 61 LMKGVLDMPFELCRRSEDIYELPLSPMYRLLY-DORDIFVYSDRE-----NM 108
Db 61 TRHVLDALFAACGTTRAEEHLTLPLRLARHFWPDGSSLDLPTDTEANTEARAFAGDK 120
QY 109 RAEIQRVFDEGTG-YTQFMQEQRKRNALYP-----CITRD--YSSIKSFLSL-DL 156
Db 121 EAAAFRRPDLTTGLWEAF--HRSVIAAPKPDILWRIAATAATVTRPQLWPAIRPGLTWKDL 177
QY 157 IKALPWLAFPKSVFNNLGOYFNQEKELAFQSKYLQMSWPCALFTMLPYLEHEYGI 216
Db 178 -----LAHFQDPLQLQRFQRTATYVGGPRGATPAVLISLIWQAEVQ-GV 220
QY 217 YHVKGGLNRIAAAOVIAENGGEIHLNSETESIENGAAGKGVKLOHGAELRGDEVLIN 276
Db 221 WAIREGMHVAALARAARVAEAKGVRFHYGAKAKRIVRKEGRTAVETVETGVSIPGACIFN 280
QY 277 ADPAHAMTHLVKGVKX-----KYTPENLQREYSCSTFWMLYGLDKI-YDLPHHTI 327
Db 281 GD-----PGALRDGLLGAARASMEKSRPAPSLSAWVAFGATPIGVDLAHNV 330
QY 328 VFAKDYTTNIRINFDNKTLDFFSYQNASADSL-APAGKSALYLV-----PM 378
Db 331 FFTADPELEFGPIGAGE-MPEEPTLYI---CAQDREMOQAPVPIERFEILMNGPAGHQP 386
QY 379 PNNSGLDQWQHCONVREQVLDITLGLSLDRAHTECKIITPOTWETDEHYKAGT 438
Db 387 PQE-----EAQC---RARTFPLAA-MGLT-FSPDPETRALTPALLSRPFGSLGAIY 435
QY 439 SLSHKFSOMLYWRPNRFEELANCYLVGCGTHPGSGLPITYESARISAKLI 489
Db 436 GGSPECTLATFRPLAR-TGLKGLYLAGSGTHPGAGVPMALTSQTHAARAL 485

RESULT 15
P49_STRLI
ID_P49_STRLI STANDARD; PRT; 469 AA.
AC P06108;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE P49 protein.
GN P49.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=66 / 1326;
RX MEDLINE=87231086; PubMed=34531116;
RA Burnett W.V., Henner J., Eckhardt T.;
RT "The nucleotide sequence of the gene coding for XP55, a major
RT secreted protein from Streptomyces lividans."
RL Nucleic Acids Res. 15:3926-3926(1987).
CC -----

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Y00142; CAA68336.1; -;  
 PIR; S09189; S09189;  
 InterPro; IPR000759; Adrndx reductase.  
 InterPro; IPR000205; NAD\_BS  
 InterPro; IPR008151; Phyt\_n\_dehydro.  
 PRINTS; PR04119; ADXRD1ASE.  
 ProDom; PD139017; Phyt\_n\_dehydro; 1.  
 SEQUENCE 469 AA; 49740 MW; C2AE7533C7C701CB CRC64;

Query Match 6.5%; Score 175; DB 1; Length 469;  
 Best Local Similarity 21.8%; Pred. No. 1.2e-05;  
 Matches 118; Conservative 70; Mismatches 221; Indels 132; Gaps 26;  
 8 IIVGAGPGICAGMILLSORGFKVSIPTDGHAEIGG--RNRPINMGPTFD-----T 55  
 5 VVVGAGPGLTAALVARRPFFVFAQGTGGGARTTEELTFGRHDPCCSAHPGAIN 64  
 56 GPTFLMKGVLDMEFELCERRSEDVLEFL-PLSPMYRLLYDORDIFVYSRENNRAELQR 114  
 65 SPAF---RGLPLERYG-----LEWLHGLPNAH-PFPGSSAAVLSRSGVETAASFG 111  
 115 VFDEGTGTYEQMEQERKFNALYPCITRDYSSLSKFSLSLILKALP-----WLA 164  
 112 ARDAGP--YRRLLIE---RFLPRWDTLARDFMSLP-----LTALPRDPVTLARFGLVG 158  
 165 FPKSVFNILGOYFNOEKMLAFCSQKVLGMSPP--WECPALFTMLPYLEHEYGIYHVKG 221  
 159 LPPSTW--LMRRFRDEKAKTLFAGLVAVH-MAPLGGFATGAIGLVFALAARHGWFWARG 215  
 222 GLNRIAAAMAQVIAENGGEIHLNSESIL-----IIENGAAGVKL-----OHGAEL 268  
 216 GSGSISDALTAYLKDLGGAVHTDYEVKRLDDLPPARAYVLDTSPTALARIAGLGSYVANY 275  
 269 RGEVNIINADPAHATHLVKGVKKYTPENLKQREYSCSTFMYLVGLDKYIDLPHHTIV 328  
 276 RYGPSVFKIDYA-----LDGP---VPWTAEEPR-----SACTVOIGAD----- 310  
 329 FAKDYTTNIRNIFDNKLTLD-----DFGFFVQNASASDDSLAPAGSKALYVLPMPNDS 383  
 311 -----STEIGALHAPSGTDRAPERFLLITQ-PSVADPTAPAGKHVFWAYGVPNG-- 362  
 384 GLDQAHQONVREQVLDTLGRLGLSDIRAHIECEKIITPQTWETDEHYKATFS----- 439  
 363 ---WNGDLTDAZEROLERTA-----PGFEDRVLARATAGPAELARNANYVGGDISGAV 414  
 440 -----LSHKFSOMLYWRPHNRFEELANCYLVGCGTHPGSGLPTIYESARISAKLISOKH 493  
 415 SGLQLLRPKISLFFPYSTEH-----PAVFICSSATPPGPGVHGM--SGHNAKAVWERL 466  
 494 R 494  
 467 R 467

earch completed: February 29, 2004, 14:44:59  
 ob time : 12.6124 secs

1 protein - protein search, using sw model

in on: February 29, 2004, 14:34:14 ; Search time 14.2234 Seconds  
(without alignments)  
3455.835 Million cell updates/sec

tie: US-09-941-947A-22

fect score: 2711

quence: 1 NANTKHIIVGAGGGLCAG.....KRVRFKDIAHSAWLKAKA 511

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

al number of hits satisfying chosen parameters: 283366

nimum DB seq length: 0

ximum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

tabase :

PIR 78:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	929.5	34.3	490	2 AH1199	phytoene dehydroge
2	915.5	33.8	488	2 AF1557	phytoene dehydroge
3	883.5	32.6	517	2 T31463	probable diaphyrt
4	825.5	30.5	529	2 S35306	phytoene dehydroge
5	794.5	29.3	499	2 A25399	zeta-carotene desa
6	794	29.3	517	2 S32169	hypothetical prote
7	783.5	28.9	499	2 S43324	zeta-carotene desa
8	743	27.4	524	2 A32617	phytoene dehydroge
9	737	27.2	518	2 S49620	phytoene dehydroge
10	735	27.1	518	2 T50745	phytoene dehydroge
11	725	26.7	511	2 T50910	phytoene dehydroge
12	713.5	26.3	492	2 S25986	phytoene dehydroge
13	704.5	26.0	492	2 A33120	phytoene dehydroge
14	702.5	25.9	502	2 B90051	squalene synthase
15	700.5	25.8	492	2 D37802	phytoene dehydroge
16	686	25.3	583	2 S43139	phytoene dehydroge
17	681	25.1	595	2 A35919	carotenoid biosynt
18	673	24.8	582	2 T46822	phytoene desaturas
19	651.5	24.0	548	2 C75466	phytoene dehydroge
20	613.5	22.6	448	2 B55458	crtn protein - Sta
21	607	22.4	621	2 T48646	phytoene dehydroge
22	596	22.0	514	2 E59108	phytoene dehydroge
23	564.5	20.8	523	2 T36968	probable phytoene
24	547.5	20.2	512	2 H84320	phytoene dehydroge
25	504	18.6	485	2 S32171	hydroxyneurosporen
26	502.5	18.5	497	2 E90061	hypothetical prote
27	485.5	17.9	536	2 B84327	phytoene dehydroge
28	472.5	17.4	528	2 T51119	phytoene desaturas
29	404	14.9	525	2 T50893	methoxyneurosporen

30	392.5	14.5	454	2 A99470	phytoene dehydroge
31	354	13.1	498	2 H83880	hypothetical prote
32	332.5	12.3	542	2 S76617	hypothetical prote
33	325	12.0	494	2 S04406	methoxyneurosporen
34	324	12.0	495	2 S21633	probable phytoene
35	323	11.9	511	2 E75561	methoxyneurosporen
36	322	11.9	486	2 S49624	hypothetical prote
37	314	11.6	565	2 A12273	methoxyneurosporen
38	290	10.7	486	2 T50749	probable carotenoi
39	278	10.3	506	2 T34971	hypothetical prote
40	276.5	10.2	587	2 A86203	hypothetical prote
41	273.5	10.1	503	2 AC2446	hypothetical prote
42	256.5	9.5	506	2 AB2064	phytoene dehydroge
43	249	9.2	518	2 G90413	hypothetical prote
44	249	9.2	574	2 A96612	hypothetical prote
45	248	9.1	507	2 S74689	hypothetical prote

ALIGNMENTS

RESULT 1

AH1199

phytoene dehydrogenase homolog lmo1000 [imported] - Listeria monocytogenes (strain EGD-e  
C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001

A:Accession: AH1199

R:Glaser, P.; Prangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AH1199

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-490 <GLA>

A:Cross-references: GB:NC 003210; PIDN:CAC99078.1; PID:G16410402; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo1000

Query Match 34.3%; Score 929.5; DB 2; Length 490;

Best Local Similarity 37.6%; Pred. No. 4.3e-62;

Matches 184; Conservative 98; Mismatches 199; Indels 9; Gaps 1;

QY	5	KHIIIVGAGGGLCAGMLLSORGFKVSIFFDKHAIIGGNNRPINNNGFTDGTFLMKG	64
DB	7	KKAIATIGAGPGGLAAGMLLSQGYQVMIYKNDRIIGRTALHRMGKYSFDFGPGSALNTW	66
QY	65	VLDENFELCERRSDYLEFFLPSPMYRLLYDDRDIFVYSDRENMRALQRFDEGTGYE	124
DB	67	VLTSLFMDCNENILDYVSLPINIHTLYPKDITFLYSDQSETKAVIQTFPGSEDFD	126
QY	125	QFMQERKRFNALYPCITRDYSSLSKSFUSLDLIKALPWLAPPKSVFNNLGOYFQNERML	184
DB	127	RFMKENTKMLYISPLNQFNYSLSLDFRPTTLRAIPSLTLGRSLMDLGRYFNKYLRL	186
QY	185	APCFOSKYLGSPECPALFTMLPVLHEHYGIHVKGGLNRIAAAMAQVLAENGGEIHLN	244
DB	187	AFSLQVRVLGSPWDIPAYSIIPFSEVYGTFTPIGGQNKIVEAMQOVVTEKNGKFFN	246
QY	245	SSIESLIIENGAAGVKVLOHGAELRGDEVIINADFAHAMTHLVKPGVLKKTYPENLKRE	304
DB	247	SEVTFESNGEITGAVLANGKTIEADYFTNLDFIYSLTN-----EHPDKLTKE	297
QY	305	YSCSPFMYLGLDKYDLPHHTIVFAKDYTTNIRNIFDNKTLTDDFSYYQNASASDLS	364
DB	298	YSSAFIDYLGKTVLPFSSHOSIIPENYREFAHNTWHKILSKDIAHLNTPSATNTM	357
QY	365	APAGKSALYVLVPMNNDSGLDWOAHCONVREQVLDLTGLARGLSLRAHIECEKIIITPQ	424





Reference number: S35306; MUID:93223667; PMID:8467787

Accession: S35306

Molecule type: DNA

Residues: 1-529 <FON>

Cross-references: EMBL:M94727; NID:g150079; PIDN:AAA25390.1; PID:g150080

Genetics:

Gene: carC

Superfamily: phytoene dehydrogenase

Keywords: carotenoid biosynthesis; oxidoreductase

Query Match 30.5%; Score 825.5; DB 2; Length 529;  
Best Local Similarity 35.1%; Pred. No. 3.3e-54;  
Matches 179; Conservative 106; Mismatches 192; Indels 33; Gaps 10;

3 NTKHIIIVGAGCGGLCAGMLLSQSGFKVSIQFKHABIGGRNRPINMNG---FTFDGPTF 59

4 KHIIVGAGCGGLCAGMLLSQSGFKVSIQFKHABIGGRNRPINMNGFTTDTGPTFLMK 63

5 KHIIVGAGCGGLCAGMLLSQSGFKVSIQFKHABIGGRNRPINMNGFTTDTGPTFLMK 64

6 LLMGVDEMFCERSEDYLFPLSPMYRL-LYDDRDIFVYSDRENMEALQRFVDEGTDG 122

62 HLYKELFEEAGLNADYVQLKRLFPYTLKFWDTQDITSDQSFKTQLATLRSDLPLA 121

123 YEQFMEOQERKFNALY-PCITRDYSSLSKSLD-LIKAL---FWLAPPKSVFNLLQYF 177

122 FDRWYSEHIRKVELGYKPYLAGPARSIFGVLRPDELKMLSFREW---ENLYOHFWRF 177

178 NOEKRLAFCQSKYLGMSPWECALFTMLPYLEHYGIYHVKGGLNRIIAAQAQVIAEN 237

178 QDRLVVALYSPSKYLGHPVAVSSVFLSIPPELFSGQVWHPVGGPRALLAQGLANAQDL 237

238 GGEIHLNSEIESLIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLYKYP 297

238 GVKHLHSPVHQIWIQGVGEGLELADASRHPQFTVINADFAVAVRHLLPPTSARGTYD 297

298 ENLKOREYSCSTFMVLYGLDKIY-DLPHHTIVFAKOTYTNIRNI---FNKTLTD-DPS 351

298 NKLQGMQPCSTFMVLYGLDKIY-DLPHHTIVFAKOTYTNIRNI---FNKTLTD-DPS 351

352 FVQNASASDDSLAPAGKSAVLYVPMNDSGLDQAHQCNVREOVLDTLGARLGLSDI 411

354 FVQNASASDDSLAPAGKSAVLYVPMNDSGLDQAHQCNVREOVLDTLGARLGLSDI 411

412 RAHIECEKIIPOTWETDEHYKATFSLSHKFSOMLYWRPHNPFELANCYLVGQTHP 471

413 EOHIVTQSCYTAQSWLDYRVHLGAVFNLSHNLTLQGFPRPPPIRSENIAGLYWTGGVHP 472

472 GSGLTIVESARISAKLSQ 491

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

473 GSGLTIVESARISAKLSQ 492

```

65 VLDEMFELCRRSEDYLEFLPLSPMYRLLYDDR-DIFVYSDRENMAELQRFVDBGTGY 123
66 VFEETFRVGRRTEDYTLRLCDPNRYRVHFRDRSVTFITSELCAWGERLVERPEGSARY 127
124 EQFMEQERKEF-NALYPCITRDYSSLSKLSFLSLDLKALPWLAPFKSVFNNLGQYFNQRM 182
128 LAFLAQGRVQYRTSLDHLVGNVAGLRDYLSPVLARIFQVRAHRMYADVSRFFODERL 187
183 RLAFCTQSKYLGMSPECPALFTMLPYLBEHYGIYHVKGGLNRIAAAMAQVIAEN 242
188 RAATFTQWYLGVSYPASPAVYGLDFTLGLVGIWFKGGLYALPOALERARREGVRFH 247
243 LNSEISLIIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLKYYTLENLKQ 302
248 YGAPVERILTDGGRTRGVRLGEGVEVEADAVLCNADLPYAVEKLLDP---KATTLKRKEK 304
303 REYSCSTFMVLYGLDKIY-DLPHHTIVFAKDYTNIRNFNKTLLTDDSPFYVQNASAD 361
305 LRYTSSGYMLYLGMRARYPELLHNVVFGDYKSGFDDIFEFPR-VPEDSPFYVNAPTRTD 363
362 DSLAPAGKSALYVLPMPNNDGLDQAHCONVREQVDTLGLARGLSLDIRAHISCE-KI 420
364 ASLAPGKDALYVLPVPHQHPDLDMKVEGPKYRAKEFARM-AELGFPLESDDIEVERIS 422
421 ITPQWETDEHYVKGATFSLSHKFSQMLYRPNRFEELANCYLVGSGTHPGSGLPTIYE 480
423 STPDMAGTENLARGSGFLGSONFTQIGFPRSPNODARVKNLFFVGASTQPGTGLPTVLI 482
481 SAR-ISAKLISQKH 493
483 SARLVTERLMTWAH 496

RESULT 7
43324
zeta-carotene desaturase - Anabaena sp. (strain PCC 7120)
;Species: Anabaena sp.
;Variety: PCC 7120
;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
;Accession: S43324; S62214
;Linden, H.; Misawa, N.; Saito, T.; Sandmann, G.
Plant Mol. Biol. 24, 369-379, 1994
;Title: A novel carotenoid biosynthesis gene coding for zeta-carotene desaturase: function and localization in Anabaena
;Reference number: S43324; MUID:94154256; PMID:8111038
;Accession: S43324
;Molecule type: DNA
;Residues: 1-499 <LIN>
;Cross-references: GB:D26095; NID:G439477; PIDN:BAA05091.1; PID:G439478
;Experimental source: PCC7120
;Albrecht, M.; Linden, H.; Sandmann, G.
Eur. J. Biochem. 236, 115-120, 1996
;Title: Biochemical characterization of purified zeta-carotene desaturase from Anabaena
;Reference number: S62214; MUID:96184887; PMID:8617254
;Accession: S62214
;Molecule type: protein
;Residues: 2-11 <ALB>
;Superfamily: phytoene dehydrogenase
;Keywords: carotenoid biosynthesis; membrane bound

Query Match 28.9%; Score 783.5; DB 2; Length 499;
Best Local Similarity 35.0%; Pred. No. 4.3e-51;
Matches 175; Conservative 103; Mismatches 201; Indels 21; Gaps 10;

2Y 4 TKHIIIVAGPGGICAGMLLSQRFKYSIFDKAEIGRNPINMGFTDGTGTFLLMK 63
2b 2 SKKVAIVAGPGGATAIRLACGLQVEIEFAERVGRNGRGEVDSYAFDTGTTILQLP 61
7Y 64 GVLDMEFELCRRSEDYLEFLPLSPMYRL-LYDDRDIFFVYSDRENMAELQRFVDBGTG 122
2b 62 HLYKELFEAGLNPAFYVQLKLESPYTLKAFWDGTQDITSDLSQSFQTLATLRSOLPLA 121
7Y 123 YEQFMEQERKEFNALY-PCITRDYSSLSKLSFLSL-DLIALKAL---PWLAPPKSVFNNLGQYF 177

```

```

122 FDRWYSEHIRKVELGYKPYLAGPARSIFCYLRPDDLMKZLSFRPW-----ENLYQHFWRFF 177
178 NQEKMLAFCTQSKYLGMSPECPALFTMLPYLBEHYGIYHVKGGLNRIAAAMAQVIAEN 237
178 QDERLVYDRLPYKYLGMHTPTVASSVSLPIFLERSQGVVHPVGGFRALAOGLANAQDL 237
238 GGEIHLNRIESLIIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLKYYTTP 297
238 GVKIHLSPVHQIWIDQGVGRGLEADASRHOFTVWINADPAYAVRHLLPTTSARGRYTD 297
298 ENLKOREYSCSTFMVLYGLDKIY-DLPHHTIVFAKDYTNIRNI---FDNKTLLTD-DPS 351
298 NKLGMQWQFSCSTFMVLYGINRRYEDLPHEQIYI-----SDNIRLERPWVDDSDALDETDP 353
352 FYVQNASASDDSLAPAGKSALYVLPMPNNDGLDQAHCONVREQVDTLGLARGLSLDI 411
354 FYVCPNPTIIDPSNAPAGHSTLFLVLEIPNTSAVOWMDIKQKSYTDIFILKRLHL-LGYENI 412
412 RAHIECEKIITPQWETDEHYVKGATFSLSHKFSQMLYRPNRFEELANCYLVGSGTHP 471
413 EQHIVTQSCYTAQSNLDDYRVHUGAVFNLSHNLTLQGLPFRPPIRSNIAGLWIGGAVHP 472
472 GSGLPITYESARISAKLISQ 491
473 GSGLLTILBASRSAGFIHQ 492

RESULT 8
A32617
phytoene dehydrogenase (EC 1.3.-.-) - Rhodobacter capsulatus
N.Alternate names: phytoene desaturase
C.Species: Rhodobacter capsulatus
C.Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 29-Sep-1999
C.Accession: A32617; S04402
R.Bartley, G.E.; Scolnik, P.A.
J. Biol. Chem. 264, 13109-13113, 1989
A.Title: Carotenoid biosynthesis in photosynthetic bacteria. Genetic characterization of
A.Reference number: A32617; MUID:89327279; PMID:2546948
A.Accession: A32617
A.Molecule type: DNA
A.Residues: 1-524 <BAR>
A.Cross-references: GB:J04969; NID:G340554; PIDN:AAA50313.1; PID:G556397
R.Armstrong, G.A.; Alberti, M.; Leach, F.; Hearst, J.E.
Mol. Gen. Genet. 216, 254-268, 1989
A.Title: Nucleotide sequence, organization, and nature of the protein products of the c
A.Reference number: S04401; MUID:89313663; PMID:2747617
A.Accession: S04402
A.Molecule type: DNA
A.Residues: 1-524 <ARM>
A.Cross-references: EMBL:X52291; NID:G45996; PIDN:CAA36533.1; PID:G45998
A.Note: translation of codons 1-33 is not given
A.Note: the authors translated the codon GTG for residue 34 as Met
C.Genetics:
A.Gene: crtI
C.Superfamily: phytoene dehydrogenase
C.Keywords: carotenoid biosynthesis; oxidoreductase

Query Match 27.4%; Score 743; DB 2; Length 524;
Best Local Similarity 32.8%; Pred. No. 5.2e-48;
Matches 162; Conservative 103; Mismatches 211; Indels 18; Gaps 9;

QY 8 IIVAGPGGICAGMLLSQRFKYSIFDKAEIGRNPINMGFTDGTGTFLLMKGVLD 67
12 VVIAGLGGLAAARLGAQGYKTVVDRLDRPGGRSSITKGGHFDLGTITVTPDLRL 71
68 EMFELCRRSEDYLEFLPLSPMYRLLYDDRDI-FVYSDRENMAELQRFVDBGTGYEQF 126
72 ELWADCGEDFDKQVSLVPMEFFITIDPDGCKYTAYGDDAKVAEVARISPGDVEGFRHF 131
127 MEQKRKFNALYPCITRDYSSLSKLSFLSL-DLIALKALP---WLAPPKSVFNNLGQYFNQEK 182
132 MWDARAYEFCYENLGR-----KFMKSLWDLIKVLPFTGWLADRVSIVYGHAKMKVDDHL 186

```



479 THPGAGVPGVIGSAEYNAKLPDAPRAR 506

## !RESULT 11

50310  
 ;Protein: phytoene dehydrogenase [imported] - Rubrivivax gelatinosus  
 ;Species: Rubrivivax gelatinosus  
 ;Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 28-Jul-2000  
 ;Accession: T50910  
 ;Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.  
 ;Submitted to the EMBL data Library, November 1999  
 ;Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosyn  
 ;Reference number: Z25270  
 ;Accession: T50910  
 ;Status: preliminary; translated from GB/EMBL/DBJ  
 ;Molecule type: DNA  
 ;Residues: 1-511 <NAG>  
 ;Cross-references: EMBL:AB034704; PIDN:BAA94063.1  
 ;Experimental source: strain IL144  
 ;Genetics:  
 ;Gene: crtI  
 ;Superfamily: phytoene dehydrogenase

Query Match.	26.7%;	Score 725;	DB 2;	Length 511;
Best local Similarity	32.4%;	Pred. No. 1.1e-46;		
Matches	158;	Conservative	98;	Mismatches 208; Indels 24; Gaps 8
8	IIVGAGGGGLCAGMLLSQRCFKVSI	FDKHAEEIGRNRPINMGFTDTGPTFLMKGVLD	67	
12	LVVSGGGGMAAVRLAAKGYRVIVLEKLD	APGGRAYVIRREGHVFDAAGTTIVVPYFLD	81	
16	EMFELCERRSEDYLEFLSPMYRLLYDDRDIFVYIS	DRENMAELQRFVDEGTDGYBQF	126	
20	ELNALAKRKSSDDIELKSLDPFYRIRFDGCHDFDYS	GDPAARMAEVRRISSPSDAGSERF	141	
24	MEQERKGFNALYPCITRDYSSL--KFLSL-DLIKALP	-----MLAPKSVFNNLGOY	176	
28	-----CYELGFTFLGDKAFDTVGDDLIKAAPLIV	KLRGW-----RSLQWVSSH	190	
32	FNQEKRLAFCFQSKYILGMSPECPALFTMLPYLE	HEHYGIYHVKGGLNRIAAANAQVIAE	236	
36	LKHPKLRIAMSLQGLLIIGGPFSSVTSMYALVNAL	ERQWGVHNAWGTTGELRGLVDVPEG	250	
40	NGGSIHILNSTRISLIIENGAAKGVKLQHGAE	LRGDEVIINADFAHAMTHLVKPGVLKXYT	296	
44	MGGMRLKAEVKRIEVDNGVATGVTLLADGERI	PADIVCNGDGTGYLKNLVDARWKHWT	310	
48	PENIKQREYSSCTMLVGLDKYI-DLPHHTIIPAK	DYTTNIRNIENDKILTDDFSFYVQ	355	
52	DARIERGHYSMGLFVWYFTGDRREYEDVPHMA	VNLGPRYRELLDDIERKKKGLASDPSIYLH	370	
56	NASASDSDLPAGKSAIYVLVPMPNDSGLDQA	HCNQVREQVLDLTGLARLGLSDIRAH	415	
60	RPTATDSMPAGACDITYALMPVPHLGS	GGTDWTTOAEPTQSVQEAL-ERTVLPLGLQHL	429	
64	ECEKIITPQWETDEHYVKATFSLSHKFSQ	MLYWRPHNFEELANCYLVGGGTHPGSL	475	
68	RVSPTCTFLDFQHLLSVKKGAGGLEP	LL--QSAFYRPNRNSVDVKLFWMGASTHPG	489	
72	PTTIESAR	483		
76	PGVIMS	497		

RESULT 12

352586  
 pyruvate dehydrogenase (EC 1.3.-.-) - Erwinia herbicola  
 Species: Erwinia herbicola  
 Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 29-Sep-1999  
 Accession: J02586  
 Lin, Y.P.; Lai, E.M.; To, K.Y.; Chang, Y.S.; Liu, S.T.  
 Mol. Gen. Genet. 245, 417-423, 1994

A,Title: Transcriptional activation of flanking sequences by Tn1000 insertion.  
A,Reference number: S52583; NUID:95107237; PMID:7308390  
A,Accession: S52586  
A,Status: preliminary; nucleic acid sequence not shown; translation not shown  
A,Molecule type: DNA  
A,Residues: 1-492 kLIN->  
A,Cross-references: EMBL:M90598; NID:G148393; PIDN:AA21263.1; PID:gl48397  
A,Note: the nucleotide sequence was submitted to the EMBL Data Library, April  
C,Superfamily: phytoene dehydrogenase  
C,Keywords: oxidoreductase

Query Match	26.3%	Score	713.5	DB	2	Length	492
Best Local Similarity	31.9%	Pred.	No.	7.9e-46			
Matches	158	Conservative	107	Mismatches	208	Indels	25
Gaps	10						
Qy	5	KHIIIVGAGGGICAGMLLSQRFQKVSIFDKHAEIGGRNRPINMNGSETFTDGTPTFLLMKG	64				
Db	2	KPTIVIGAGFGGLAIRLQAAGIPVLLLRQDKPGRAVYVSDQGSFTDAGPTVITDPS	61				
Qy	65	VLDMPFLCERRSEDYLFELPLSPMYELLYDDRDDIFVY-SDRENWRAELORVFECDTGY	123				
Db	62	AIBELFTLAGKQKQVYVELLPVAPFFKLCWESGKVFYNDNDQLEAQIQQFNPRDVEGY	121				
Qy	124	EQPMEOERKTFNALYPCITRDYSSLK-----SFLSL-DLIALPWLA---FPKSVPNNLQG	175				
Db	122	RQFLDYSSRAVFKSGY-----LKLGTVPFLSFRDMLRAAPQLAKLQAWRTVYSKVAS	172				
Qy	176	YFNOEKRLAFCTQSKYLGMSPHECPALFMLPVLYHEHYGIYHVKGGLNRITAAAMAQVIA	235				
Db	173	YIEDEHLRQAFSHSLVGNPGTATSIYTLIHALEREWGVPFRGTGALVKGMKILQF	232				
Qy	236	ENGGBIHLNBEISLIIENGAAKVKLQHGAELRGDEVIINADFAHAMTHLVK--PGVLK	293				
Db	233	DLGGEVYLNAAKSHMETGTGTIEAVHLEDGRDPPETRAVASNADVVHTYRDLSSQHPAAVK	292				
Qy	294	KYTPENLKOREYSCSTNMLVLGLDKIYD-1PHHTIVAKDYTTNIRIENKTLTDDPSF	352				
Db	293	Q--SKGLQTRMNSLFVLYFGLNHHHDQLAHTVCFGPRYRELIHEIFNHDGLADDFSL	350				
Qy	353	YVQNASASDDSLAPAGKASALYVLYVPMEN-NDSGLDWQAHCONVREQVLDITGARGLSDI	411				
Db	351	YLHAPCVTDSLAPBEGCGSYVLAPVPHLGTALNDWTVEGPRLDRIFFEYLEQHY-MPLG	409				
Qy	412	RAHTECKLIPTQWETDEHYKATSLSHKKSOMLYMRPHNRFEEALNCYLVGCGGTHP	471				
Db	410	RSQIVTORMFTTFDRQLNAYQGSASFVEPVLQSAWFRPHNRDKTINNLIYLVAGGTHP	469				
Qy	472	GSGLPTTVEGARISAKLI	489				
Db	470	GAGTPEGVTGSAKATAGLM	487				

RESULT 13

A33120  
phytoene dehydrogenase (EC 1.3.1.1) - Erwinia herbicola  
N:Alternate names: phytoene desaturase  
C:Species: Erwinia herbicola  
C:Date: 28-Sep-1990 #sequence\_revision 28-Sep-1990 #text\_change 29-Sep-1999  
C:Accession: A3273; A33120  
C:Accession: A3273; A33120  
R:Armstrong, G.A.; Alberti, M.; Hearst, J.E.  
Proc. Natl. Acad. Sci. U.S.A. 87, 9975-9979, 1990  
A:Title: Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nor  
A:Reference number: A3273; PMID:91088634; PMID:2263648  
A:Accession: A3273  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-492 <REM>  
A:Cross-references: GB:K8423; NID:G148401; PIDN:AAA24820.1; PID:G148402  
A:Note: the authors translated the codon CAG for residue 181 as Phe, TCC for residue 181  
C:Genetics:  
A:Note: strain Eho10; ATCC 39368  
A:Gene: crtI  
C:Superfamily: phytoene dehydrogenase



db 233 DLGGEVILNARVSHMETTGNGKIEAVHLEDGRRFLTQAVASNADVVHTYRDLLSQHFAAVK 292  
2y 294 KYTENLKOREYSGSTFWLYGLDKIYD-LPHHTIVFAKDYTTNIRNIPDNKTLTDDFSP 352  
db 293 Q--SNKLOTKMSNSLFVYFGLNHHDDQLAHTVCFGPRIYRELIDEIFNHDGLAEDFSL 350  
2y 353 YVQNASASDDSLAPAGKSALYVLYVMPN-NDSGLDQAHCONVREQVLDTLGARLGLSDI 411  
db 351 YLHAPCVTDSSLAPEGCGSYVVLAPVPHLGTANLDWTEGPKLRDRIFAVLEQHY-MPGL 409  
2y 412 RAHIECEKIITPQWETDEHVYKGAATESLSHKFSOMLYWPHNRFELANCYLVGGGTHP 471  
db 410 RSQLVTHRMFTFPDFRDQANAYGSAFSEVPEVLTQSAWFRPHNRDKTITNLYLVGAGTHP 469  
2y 472 GSGLPTIYESARISAKLI 489  
db 470 GAGIPGVIGSAKATAGLM 487

Search completed: February 29, 2004, 14:52:39  
Job time : 17.2234 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

WM protein - protein search, using sw model

Run on: February 29, 2004, 14:51:24 ; Search time 33.9275 Seconds  
(without alignments)  
3180.293 Million cell updates/sec

Title: US-09-941-947A-22

Perfect score: 2711

Sequence: 1 MANTKHIIVGAPGGLCAG.....KRVRFKDIASHANLKAKA 511

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/prodata/2/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/prodata/2/pubaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/prodata/2/pubaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/2/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/prodata/2/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/2/pubaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/prodata/2/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/prodata/2/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/prodata/2/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/prodata/2/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/prodata/2/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/prodata/2/pubaa/US09D\_PUBCOMB.pep.\*
- 13: /cgn2\_6/prodata/2/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/prodata/2/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/prodata/2/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/prodata/2/pubaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/prodata/2/pubaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/prodata/2/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	2711	100.0	511	9	US-09-934-903-16
2	2711	100.0	511	9	US-09-934-868-76
3	2711	100.0	511	10	US-09-941-947A-22
4	2711	100.0	511	14	US-10-358-917-4
5	832.5	30.7	494	15	US-10-369-493-18519
6	816	30.1	517	15	US-10-369-493-19519
7	794.5	29.3	499	15	US-10-369-493-18983
8	767	28.3	498	15	US-10-369-493-20438
9	740	27.3	518	15	US-10-369-493-7750
10	703.5	25.9	492	10	US-09-941-947A-32
11	703.5	25.9	492	14	US-10-218-118-8
12	702.5	25.9	502	14	US-10-358-917-12
13	699.5	25.8	491	15	US-10-369-493-234
14	651.5	24.0	548	15	US-10-369-493-540
15	640.5	23.6	494	9	US-09-547-267-5

16	640.5	23.6	494	10	US-09-920-923-4	Sequence 4, Appli
17	639.5	23.6	498	15	US-10-369-493-19248	Sequence 19248, A
18	638.5	23.6	618	15	US-10-369-493-3769	Sequence 3769, Ap
19	634.5	23.4	498	15	US-10-369-493-10579	Sequence 10579, A
20	627.5	23.1	404	15	US-10-369-493-19547	Sequence 19547, A
21	609.5	22.5	506	15	US-10-369-493-17930	Sequence 17930, A
22	607.5	22.4	497	9	US-09-934-903-18	Sequence 18, Appli
23	607.5	22.4	497	9	US-09-934-868-78	Sequence 78, Appli
24	607.5	22.4	497	10	US-09-941-947A-24	Sequence 24, Appli
25	607.5	22.4	497	14	US-10-358-917-8	Sequence 8, Appli
26	596	22.0	514	15	US-10-369-493-21541	Sequence 21541, A
27	571.5	20.1	506	15	US-10-369-493-8142	Sequence 8142, Ap
28	547.5	20.2	512	15	US-10-369-493-18644	Sequence 18644, A
29	521	19.2	513	14	US-10-156-763-8562	Sequence 8562, Ap
30	501	18.5	485	15	US-10-369-493-19518	Sequence 19518, A
31	500.5	18.5	497	14	US-10-358-917-14	Sequence 14, Appli
32	485.5	17.9	536	15	US-10-369-493-18646	Sequence 18646, A
33	480	17.7	548	9	US-09-738-626-4194	Sequence 4194, Ap
34	470	17.3	1268	15	US-10-438-784-3	Sequence 3, Appli
35	467	17.2	498	15	US-10-369-493-10196	Sequence 10196, A
36	451	16.6	496	15	US-10-369-493-20440	Sequence 20440, A
37	447.5	16.5	439	9	US-09-815-242-5696	Sequence 5696, Ap
38	423.5	15.6	469	15	US-10-438-784-5	Sequence 5, Appli
39	409.5	15.1	530	14	US-10-128-713A-18	Sequence 18, Appli
40	354	13.1	498	15	US-10-369-493-17259	Sequence 17259, A
41	332.5	12.3	542	14	US-10-209-372-6	Sequence 6, Appli
42	332.5	12.3	542	15	US-10-369-493-2831	Sequence 2831, Ap
43	330	12.2	495	15	US-10-369-493-7895	Sequence 7895, Ap
44	323	11.9	511	14	US-10-209-372-4	Sequence 4, Appli
45	323	11.9	511	15	US-10-369-493-23397	Sequence 23397, A

#### ALIGNMENTS

##### RESULT 1

US-09-934-903-16  
; Sequence 16, Application US/09934903  
; Patent No. US20020102690A1  
; GENERAL INFORMATION:  
; APPLICANT: Kofas, Mattheos  
; APPLICANT: Odom, J. Martin  
; APPLICANT: Schenzle, Andreas J.  
; APPLICANT: No. US20020102690A1ton, Kelley C.  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Rouviere, Pierre  
; APPLICANT: Picataggio, Stephen  
; APPLICANT: Cheng, Qiong  
; TITLE OF INVENTION: Genes Involved in Isoprenoid Cc pounds Production  
; FILE REFERENCE: C11646 US NA  
; CURRENT APPLICATION NUMBER: US/09/934,903  
; CURRENT FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/229,907  
; PRIOR FILING DATE: September 1, 2001  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Methylobionas 16a  
; FEATURE:  
; OTHER INFORMATION: Amino acid sequences encoded by ORF8  
US-09-934-903-16

Query Match 100.0%; Score 2711; DB 9; Length 511;  
Best Local Similarity 100.0%; Pred. No. 3.8e-270;  
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MANTKHIIVGAPGGLCAGMLLSQGFVKVSIIDKHAIG ANRPINMGFTFDGPTPL 60  
Db 1 MANTKHIIVGAPGGLCAGMLLSQGFVKVSIIDKHAIG ANRPINMGFTFDGPTPL 60  
Qy 61 LMKGVLDENFELCERRSELYLEFLPLSPMYRLLYDDRDIF VSDRENMAELQRYVDECT 120



Db 61 LKMGVLDENFELCERSDYLEFLPLSPMYRLLYDDRDIFVYSRENRAELQRFVDEGT 120  
Qy 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180  
Db 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180  
Qy 181 KRLAFCFQSKYLGMSFWPCPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240  
Db 181 KRLAFCFQSKYLGMSFWPCPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240  
Qy 241 IHLNSIESLIIENGAAKGVKLGAEALRGDEVIINADFAHATHLVKPGVLYKKTYPENL 300  
Db 241 IHLNSIESLIIENGAAKGVKLGAEALRGDEVIINADFAHATHLVKPGVLYKKTYPENL 300  
Qy 301 KOREYSCSTFMYLGLDKIYDLPHHTIVPAKDYTTNIRNIFDNKLTDDFSFYQNASAS 360  
Db 301 KOREYSCSTFMYLGLDKIYDLPHHTIVPAKDYTTNIRNIFDNKLTDDFSFYQNASAS 360  
Qy 361 DDLAPAGKSALYLVPMNNDSGLDQAHCONVREQVLTGLARGLSDIRAHIECEKI 420  
Db 361 DDLAPAGKSALYLVPMNNDSGLDQAHCONVREQVLTGLARGLSDIRAHIECEKI 420  
Qy 421 ITPQWETDEHYKGAFTSLSHKFSQMLYWRPHNRFEBELJCYLVGGGTHPGSGLPTIYE 480  
Db 421 ITPQWETDEHYKGAFTSLSHKFSQMLYWRPHNRFEBELJCYLVGGGTHPGSGLPTIYE 480  
Qy 481 SARISAKLSQKRVRFKDIASHAWLKAKA 511  
Db 481 SARISAKLSQKRVRFKDIASHAWLKAKA 511

## RESULT 2

US-09-934-868-76  
; Sequence 76, Application US/09934868  
; Patent No. US20020137190A1  
; GENERAL INFORMATION:  
; APPLICANT: Koffas, Matteo  
; APPLICANT: Odom, James M  
; APPLICANT: Schenzle, Andreas J  
; TITLE OF INVENTION: IDENTIFYING METHANOTROPHIC BACTERIAL STRAIN  
; FILE REFERENCE: CL1596 US NA  
; CURRENT APPLICATION NUMBER: US/09/934,868  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/229,858  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 81  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 76  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Methylobacter 16a  
; FEATURE:  
; OTHER INFORMATION: Amino acid sequences encoded by CRTN1

US-09-934-868-76

Query Match 100.0%; Score 2711; DB 9; Length 511;  
Best Local Similarity 100.0%; Pred. No. 3.8e-270;  
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MANTKHIIIVGAGPGGLCAGMLLSQKGFVSIIPDKHAEIGGRNRPINNGGTFDTGTPTFL 60  
Db 1 MANTKHIIIVGAGPGGLCAGMLLSQKGFVSIIPDKHAEIGGRNRPINNGGTFDTGTPTFL 60  
Qy 61 LKMGVLDENFELCERSDYLEFLPLSPMYRLLYDDRDIFVYSRENRAELQRFVDEGT 120  
Db 61 LKMGVLDENFELCERSDYLEFLPLSPMYRLLYDDRDIFVYSRENRAELQRFVDEGT 120  
Qy 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180  
Db 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180  
Qy 181 KRLAFCFQSKYLGMSFWPCPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240

Db 181 KRLAFCFQSKYLGMSFWPCPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240  
Qy 241 IHLNSIESLIIENGAAKGVKLGAEALRGDEVIINADFAHATHLVKPGVLYKKTYPENL 300  
Db 241 IHLNSIESLIIENGAAKGVKLGAEALRGDEVIINADFAHATHLVKPGVLYKKTYPENL 300  
Qy 301 KOREYSCSTFMYLGLDKIYDLPHHTIVPAKDYTTNIRNIFDNKLTDDFSFYQNASAS 360  
Db 301 KOREYSCSTFMYLGLDKIYDLPHHTIVPAKDYTTNIRNIFDNKLTDDFSFYQNASAS 360  
Qy 361 DDLAPAGKSALYLVPMNNDSGLDQAHCONVREQVLTGLARGLSDIRAHIECEKI 420  
Db 361 DDLAPAGKSALYLVPMNNDSGLDQAHCONVREQVLTGLARGLSDIRAHIECEKI 420  
Qy 421 ITPQWETDEHYKGAFTSLSHKFSQMLYWRPHNRFEBELJCYLVGGGTHPGSGLPTIYE 480  
Db 421 ITPQWETDEHYKGAFTSLSHKFSQMLYWRPHNRFEBELJCYLVGGGTHPGSGLPTIYE 480  
Qy 481 SARISAKLSQKRVRFKDIASHAWLKAKA 511  
Db 481 SARISAKLSQKRVRFKDIASHAWLKAKA 511

## RESULT 3

US-09-941-947A-22  
; Sequence 22, Application US/09941947A  
; Publication No. US20030003528A1  
; GENERAL INFORMATION:  
; APPLICANT: Szostowicz, Patricia C.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: DiCossimo, Deana J.  
; APPLICANT: Koffas, Matteo  
; APPLICANT: Miller, Edward S. Jr.  
; APPLICANT: Odom, J. Martin  
; APPLICANT: Picataggio, Steve  
; APPLICANT: Rouviere, Pierre E.  
; TITLE OF INVENTION: CARBON DIOXIDE PRODUCTION FROM A SINGLE CARBON SOURCE  
; FILE REFERENCE: CL1903 US NA  
; CURRENT APPLICATION NUMBER: US/09/941,947A  
; PRIOR FILING DATE: 2001-09-01  
; PRIOR APPLICATION NUMBER: 60/229,907  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,858  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 22  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Methylobacter 16a  
; FEATURE:  
; OTHER INFORMATION: Amino acid sequences encoded by CRTN1

US-09-941-947A-22

Query Match 100.0%; Score 2711; DB 10; Length 511;  
Best Local Similarity 100.0%; Pred. No. 3.8e-270;  
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MANTKHIIIVGAGPGGLCAGMLLSQKGFVSIIPDKHAEIGGRNRPINNGGTFDTGTPTFL 60  
Db 1 MANTKHIIIVGAGPGGLCAGMLLSQKGFVSIIPDKHAEIGGRNRPINNGGTFDTGTPTFL 60  
Qy 61 LKMGVLDENFELCERSDYLEFLPLSPMYRLLYDDRDIFVYSRENRAELQRFVDEGT 120  
Db 61 LKMGVLDENFELCERSDYLEFLPLSPMYRLLYDDRDIFVYSRENRAELQRFVDEGT 120  
Qy 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180  
Db 121 DGYEQFMEQERKFNALYPCITRDYSSLSKSFSLDLIKALPWLAPFKSVFNNGQYFNQE 180  
Qy 181 KRLAFCFQSKYLGMSFWPCPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240  
Db 181 KRLAFCFQSKYLGMSFWPCPALFTWLPYLEHEYGIVHVGKGLNRIAAAMAQVIAENGGE 240

241 IHLNSESIIENGAAKGVKLOHGAELRGDEVINADFAHATHLVKPGVLKKTYPENL 300  
241 IHLNSESIIENGAAKGVKLOHGAELRGDEVINADFAHATHLVKPGVLKKTYPENL 300  
301 KOREYSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNFNDKTLTDDPSFFVQNASAS 360  
301 KOREYSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNFNDKTLTDDPSFFVQNASAS 360  
361 DDSLAPAGKSALYVLVMPNNDGLDQAHCONVREOVLTGLGARLGLSDIRAHICEKI 420  
361 DDSLAPAGKSALYVLVMPNNDGLDQAHCONVREOVLTGLGARLGLSDIRAHICEKI 420  
421 ITPQWETDEHYKATFSLSHKFSQMLYWRPHNRPELA CYLVGGTHPGSGLPTIYE 480  
421 ITPQWETDEHYKATFSLSHKFSQMLYWRPHNRPELA CYLVGGTHPGSGLPTIYE 480  
481 SARISAKLISOKHVRPKDIAHSAWLKAKA 511  
481 SARISAKLISOKHVRPKDIAHSAWLKAKA 511

RESULT 4  
US-10-358-917-4  
Sequence 4, Application US/10358917  
Publication No. US20030182687A1  
GENERAL INFORMATION:  
APPLICANT: Cheng, Qiong  
APPLICANT: No. US20030182687A1ton, Kelley C.  
APPLICANT: Tao, Luan  
TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOID COMPOUNDS  
FILE REFERENCES: C11929 US NA  
CURRENT APPLICATION NUMBER: US/10/358,917  
CURRENT FILING DATE: 2003-02-05  
PRIOR APPLICATION NUMBER: 60/355,939  
PRIOR FILING DATE: 2002-02-11  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 4  
LENGTH: 511  
TYPE: PRT  
ORGANISM: Methylobionas sp.16a  
S-10-358-917-4

Query Match 100.0%; Score 2711; DB 14; Length 511;  
Best Local Similarity 100.0%; Pred. No. 3,8e-270;  
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MANTKHIIIVGAGPGGLCAGMLLSQRFKVSIFDKHAEIGGGRNRPINMNGPTFDGTPTFL 60  
1 MANTKHIIIVGAGPGGLCAGMLLSQRFKVSIFDKHAEIGGGRNRPINMNGPTFDGTPTFL 60  
61 LMKGVLDENFELCERSESDYLEFLPLSPMYRLLYDDRDIFVYSRENKRAELQRFDEGT 120  
61 LMKGVLDENFELCERSESDYLEFLPLSPMYRLLYDDRDIFVYSRENKRAELQRFDEGT 120  
121 DGYEQMEQERKFNALYPCITRDYSSLSKSPSLDLIKALPWLAPPKSVFNNLGOYFNQE 180  
121 DGYEQMEQERKFNALYPCITRDYSSLSKSPSLDLIKALPWLAPPKSVFNNLGOYFNQE 180  
181 KMLAFCFQSKYLGMSWPCPALFTMLPYLHEHYGIYHVKGGLNRIAAQAQVIAENGGE 240  
181 KMLAFCFQSKYLGMSWPCPALFTMLPYLHEHYGIYHVKGGLNRIAAQAQVIAENGGE 240  
241 IHLNSESIIENGAAKGVKLOHGAELRGDEVINADFAHATHLVKPGVLKKTYPENL 300  
241 IHLNSESIIENGAAKGVKLOHGAELRGDEVINADFAHATHLVKPGVLKKTYPENL 300  
301 KOREYSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNFNDKTLTDDPSFFVQNASAS 360  
301 KOREYSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNFNDKTLTDDPSFFVQNASAS 360  
361 DDSLAPAGKSALYVLVMPNNDGLDQAHCONVREOVLTGLGARLGLSDIRAHICEKI 420

Db 361 DDSLAPAGKSALYVLVMPNNDGLDQAHCONVREOVLTGLGARLGLSDIRAHICEKI 420  
Qy 421 ITPQWETDEHYKATFSLSHKFSQMLYWRPHNRPELA CYLVGGTHPGSGLPTIYE 480  
Db 421 ITPQWETDEHYKATFSLSHKFSQMLYWRPHNRPELA CYLVGGTHPGSGLPTIYE 480  
Qy 481 SARISAKLISOKHVRPKDIAHSAWLKAKA 511  
Db 481 SARISAKLISOKHVRPKDIAHSAWLKAKA 511

RESULT 5  
US-10-369-493-8951  
Sequence 8951, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEIN IN PLANTS FOR PRODUCTION OF  
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
FILE REFERENCES: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 8951  
LENGTH: 494  
TYPE: PRT  
ORGANISM: Chloroflexus aurantiacus  
US-10-369-493-8951

Query Match 30.7%; Score 832.5; DB 15; Length 494;  
Best Local Similarity 36.9%; Pred. No. 1.6e-76;  
Matches 185; Conservative 97; Mismatches 189; Indels 31; Gaps 12;

Qy 7 IIVGAGPGGLCAGMLLSQRFKVSIFDKHAEIGGGRNRPINMNGPTFDGTPTFLMKGVYL 66  
Db 5 IIVGAGPGGMATAIRLAGQGYQVEIFRAVDRPGGRMGF LGDYHFDGTPTILOVPRVY 64  
Qy 67 DEMFELCERSESDYLEFLPLSPMYRL-LYDDRDIFVYSDE NMRALQRFDEGTG-YE 124  
Db 65 DELFSSAGLRFSYVTLIRLDPNTRIRFDWGEYLDLTSNI AFKAQLAR-FDPALPAEPE 123  
Qy 125 Q-FMEQERKFNALYPCITRDYSSLSKSPSLDLIKALPWLAPPKSVFNNLGOYFNQ 179  
Db 124 RWIEHIRKXNVGYEYLGTPVRSPLVYLKPREIAALAF PW----ESLYDHFRRFRD 179  
Qy 180 EKMRLAFQSKYLGMSWPCPALFTMLPYLHEHYGIYHV GGLNRIAAQAQVIAENGGE 239  
Db 180 ERYVYALSYQAKYLGMPHPTACSSVFSVLTLEFADGIVHP GPRALAAGLAKAATDLGV 239  
Qy 240 EYHNLSEIIENGAAKGVKLOHGAELRGDEVINADFAHATHLVKPGVLKKTYPEN 299  
Db 240 VHYNSFVQVLEGRACGVELGGERIKADAVVYNADF EALTHIIPPHARGYTPOK 299  
Qy 300 LKOREYSCSTFMYLGLDKIYD-LPHHTIVFAKDYTNIR- ----IPDNKTLTDD 349  
Db 300 LNSMEFSCSTFMYLGVNRWDLPFHQYL- ----SANIR NDPLWAKSAVILDE ---DD 351  
Qy 350 PSFYVQNASASDDSLAPAGKSALYVLVMPNNDGLDQAH CONVREOVLTGLGARLGLS 409  
Db 352 PSFYVQNPTVDPTNAPGHSTLFLVFPNLRVDPVDWAA EORYEDLIIRQM-AKLAGFE 410  
Qy 410 DIRAHICEKIITPQWETDEHYKATFSLSHKFSQMLY RPHNRPEELANCYLVGGGT 469  
Db 411 DVERHIVVERRYTAETWDEHYTYLGAVALNLSWSQLGP PHIRHNGARNLYWIGAV 470  
Qy 470 HPGSGLPTIYESARISAKLISQ 491

Db 471 HPGSGLMTLEAAKSAVHFIQE 492

RESULT 6

US-10-369-493-19519

Sequence 19519, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 19519

LENGTH: 517

TYPES: PRT

ORGANISM: Myxococcus xanthus

US-10-369-493-19519

Query Match 30.1%; Score 816; DB 15; Length 517;

Best Local Similarity 35.7%; Pred. No. 8.4e-75;

Matches 176; Conservative 106; Mismatches 203; Indels 8; Gaps 6;

2Y 5 KHIIVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRPINNMGFTFTDGTFTLLMK 64

Db 8 RRIIVGAGVGLAAARLAAQGFQVFXETQGPGRNRLQVQDGLGPTIVLMP 67

2Y 65 VLDEMFCERSESDYLEFLPSPMYRL-LYDDRDIFVSDRENMAELQRFDEGTDG 123

Db 68 VFEETFRVAGRIEDYTLLECDPNYVHFRGSDVFTSELCAHRELSERVEPGSY 127

2Y 124 EQFMEQERKRF-NALYPCITRDYSSLSKSLDILIKALPWLAPPKSVFNNLQYF 182

Db 128 LAFLAQGRVQYRTSLDLHVGNYAGRLDYLSPVLARIQVRAHRRMYADVSPFQDER 187

2Y 183 RLAFQFSQKYLGMSPWECALFTMLPYLEHYGIVHVGKGLNRIAAAMAQVIAENG 242

Db 188 RAATFTQTYLGVSPYASPAVYGLLPTELGVGIFPFKGLGYAIPOALERLAREEG 247

2Y 243 LNSHIESLIIENGAAGVKLQGAELRGDEVIINADFAHAMTHLVKPGVLKXYTP 302

Db 248 YGAPVERILTDGTRGVRLGEGVEADAVLGNADLPYAEKLLDP---KATLLRKEK 304

2Y 303 REYSCSTFMLYGLDKIY-DLPHTTIVPAKYDTTIRNI-FDNKLTLDDFSFFYVONAS 361

Db 305 LRYTSSGYMLYGMKRRYPPELLHNHNVFGRDYKSGPDIFERFRVPEDFSFYVNA 364

2Y 362 DSLAPAGKSALYLVPMNDSGLDQWAKCONVREQVLDLGLARGLSDIRAHIECEK 421

Db 365 ASLAPEGKDALYLVLPVPHQHPDLDMKVGKVKRAKFFARN-ALGPPLESIEVERV 423

2Y 422 TPQWTBDEHYVKGATFSLSHKFSQMLYRPHNRFEELANCYLVGGTTPGSGLP 481

Db 424 TPDDWAGTFNLARGSAFGLSQNFQIGFFRFSNQDARVKNLFFVGASTDGTGLPT 483

2Y 482 AR-ISAKLISOKH 493

Db 484 ARLVTERLATWAAH 496

RESULT 7

US-10-369-493-18983

Sequence 18983, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 18983

LENGTH: 499

TYPES: PRT

ORGANISM: Anabaena PCC7120

US-10-369-493-18983

Query Match 29.3%; Score 794.5; DB 15; Length 499;

Best Local Similarity 35.4%; Pred. No. 1.3e-72;

Matches 177; Conservative 102; Mismatches 200; Indels 21; Gaps 10;

QY 4 TKHIIVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRPINNMGFTFTDGTFTLLMK 63

Db 2 SKKVAIVGAGPGGLATAIRLAGLVQVEIFEAAERVGGRGFEVDSVAPDTGTILQLP 61

QY 64 GVLDEMFCERSESDYLEFLPSPMYRL-LYDDRDIFVSDRENMAELQRFDEGTDG 122

Db 62 HLYKELFEEAGLNFADYVQLARLEFYTLKFWDTGTDITLDQSFKTQLATERSDLPLA 121

QY 123 YEQFMEQERKRFNALY-PCITRDYSSLSKSLD-LIKA---PWLAPPKSVFNNLQYF 177

Db 122 FDRVYSEHIRKVELGYKPYLAGPARSIFGYLRPDLAMKFSEFRW---ENLYQHFWREFF 177

QY 178 NOEKLRLAFQFSQKYLGMSPWECALFTMLPYLEHYGIVHVGKGLNRIAAAMAQVIAEN 237

Db 178 QDERLVYALSPYSKYLGHPHTVASSVSFLIPFLFBSQGVHPVGGFALAGLANAAQDL 237

QY 238 GGEIHNSIESLIIENGAAGVKLQGAELRGDEVIINADFAHAMTHLVKPGVLKXYTP 297

Db 238 GVKHLHSPVQIWIHQGVRLGELADASHRQPTVVINDFAYAVRHLLPSTNARGYTD 297

QY 298 ENLKQREYSCSTFMLYGLDKIY-DLPHTTIVPAKYDTTIRNI-FDNKLTLDDFS 351

Db 298 NKLQWQFSCSTFMLYGLNRRYEDLPHQIYL----SDIRLERPWVDDSDALDETDP 353

QY 352 FYVONASASDLSLAPAGKSALYLVPMNDSGLDQWAKCONVREQVLDLGLARGLSDI 411

Db 354 FYVONPTIIDSNAPAGHSLFLVLPINTSYAVDMDIKKSTYDFILKRLHL-LGHNI 412

QY 412 RAHIECEKLIPTQWTBDEHYVKGATFSLSHKFSQMLYRPHNRFEELANCYLVGGTTP 471

Db 413 BOHIVTQSCYTAQSWLDDYRVHLGAVENLSHNLTLQGPFPPIRSENIAGLYVIGGAVHP 472

QY 472 GSGLPITYESARISAKLISQ 491

Db 473 GSGLTILEASRSAAAGFIHQ 492

RESULT 8

US-10-369-493-20438

Sequence 20438, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

```

; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-7750

Query Match 27.3%; Score 740; DB 15; length 518;
Best Local Similarity 33.7%; Pred. No. 5.8e-67;
Matches 171; Conservative 96; Mismatches 207; Indels 34; Gaps 10;

QY 8 IIVGAGPGGLCAGMLLSORGFKVISIFDKHAEIIGRRNRPINANGFTEDTDTPTFLIMKGVLD 67
Db 14 LVIGSGLGLLAAMRLGAKGWRVTVIDKLDVPGGRGSSII EGHREFDLPTITVTPQSLR 73
QY 68 EMPELCERSESDYLEPLPLSPMVELLYDDRDIF-VYSRE MEAEIQRVFDEGTGVEQF 126
Db 74 DLWKTCGRDPDADVELKPIDPFVEVWPQSGSHFTVRQSTE MKAEEVARLSPGDVAGYEKF 133
QY 127 MEQERRRFRNALYPCITRDYSYSLKSLFLSL-DLIKALP---W APPKSVFNILGOYFNQCKM 182
Db 134 LKDSERYNFGYEDLGR-----RSMKELNDLIKVLPTFCN RADRSVYQHAALRVKDERL 188
QY 183 RLAFQSQSYKLGMPWECALFTMLPYLBEHVGIVHVKGG NRIAAAMQVIAKNGGEIH 242
Db 189 RMALSPHPLFIGGDPFNVTSMILVSQLEKFGVHYAIGC AATAAAMARVIEGQGGSSFR 248
QY 243 LNSEIESLIIENCAAGKVLOHGAEURGDVEIINADFAHA THLVKPGVGLKKYTPENLQ 302
Db 249 MTEVDEIIVKEGTATGVRLASGEVLRAGLVVSNADAGHT MLLNHPRRRTDAHKVS 308
QY 303 RYSCSTFMLYLGLDKIY---DLPHHTIVPAKDYTNIE IFDNKLTLDDFSFFYQNAS 358
Db 309 RRMSMGLFVYFETGKTGMWPDVGHHTIVNAPRYKGLVE IFLKGKLAQMSLYIHRPS 368
QY 359 ASDDSLAPAGKALYVLVMPN--NDSCGLDWAHQCNVRE VLDTL-----GARLGLS 409
Db 369 IDTPTVAPEGDDTTFALSPVHLKQAQPDWQVAEPYRE VLEVEQSMNPGIGERIGPS 428
QY 410 DIRAHTECEKIIPTQWETDEHYK-GATFSLSHKFSQMI WRPHNRPBELANCYLAVGG 468
Db 429 -----LVFTPEFR-DRYLSPWAGGFSIEPRILQSAIFRPHNISEEVANLPLVAG 478
QY 469 THPGSGLPTIYESARISAKLISQKHVR 496
Db 479 THPGAGVPGVIGSAEVMKALPADAPR 506

RESULT 10
US-09-941-947A-32
; Sequence 32, Application US/09941947A
; Publication NO. US20030003528A1
; GENERAL INFORMATION:
; APPLICANT: Brzostowicz, Patricia C.
; APPLICANT: Cheng, Qiong
; APPLICANT: DiCosimo, Deana J.
; APPLICANT: Koffas, Mattheos
; APPLICANT: Miller, Edward S. Jr.
; APPLICANT: Odum, J. Martin
; APPLICANT: Pocataggio, Steve
; APPLICANT: Rouviere, Pierre E.
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE
; FILE REFERENCE: CL1903 US NA
; CURRENT APPLICATION NUMBER: US/09/941,947A
; CURRENT FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: 60/229,907
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/229,858
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 492
; TYPE: PRF
; ORGANISM: Pantoea stewartii
US-09-941-947A-32

```

Query Match 25.9%; Score 703.5; DB 10; Length 492;  
 Best Local Similarity 31.7%; Pred. No. 3.1e-63;  
 Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

QY 5 KHIIVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRNRPINMGFTFTDGTFFLLMKG 64  
 DB 2 KPTTVIGAGGGLALAIRLQAAGIPVLLLEQDRKPGKRAYVYQGGFTFDAGPTVITDPS 61  
 QY 65 VLDEMFCERRSBDYLFPLSPMYRLLYDDRDIFVY-SDENNRRAELQRFVDEGTDGY 123  
 DB 62 ALIELFALAGLQKDYVELLPVTFYRLCWESGKVFNYDNDQALEAQIOQFNPRDVAGY 121  
 QY 124 EQFMQERKRNFALPCITRDYSSLK-----SFLSL-DLI:ALPWL:---FPKSVFNNLQ 175  
 DB 122 RAFLDYSRAVNEG------LKLGTVPFLSKDMLRAAPQAKLQAWRSVYSKVG 172  
 QY 176 YFNQKMLAFQFQSKYLGMSWPCPALFTMLPYLEHEYYIYHVKGLNRIAAQAQVIA 235  
 DB 173 YIEDEHLAQAFSFLSLVGGNPPATSSIIYTLIHALEREWVMPRGCTGALVNGMKLFQ 232  
 QY 236 ENGGEIHLNSIESLIIENGAAGKVKLGHAELRGDEVIINADPAHAMTHLVK--PGVLK 293  
 DB 233 DLGGEVNLARVSHMETVGDKIQAQVLEDGRFRFTCAVSNADVVHYRDLLSQHPAAK 292  
 QY 294 KYTPENLQREYSCSTFMYLGLDKIYD-LPHHTIVFAKYTTNIRNIPFNKTLTDDFSP 352  
 DB 293 Q--AKKLQSKMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHFHHDGLAEDFSL 350  
 QY 353 YVQNASASDDSLAPAGKSAVYLVMPN-NDSGLDQAHQNVREOVLDTLGARLGLSDI 411  
 DB 351 YLHAPCVTDPSLAPBGCGSYVYLVAPVPHLGTANLDMAVEGRLDRIFDYLEQHY-MPGL 409  
 QY 412 RAHIECEKIIPTQWETDEHYKATFSLSHKFSQMLYWRPHNRFELANCYLVGCGTHP 471  
 DB 410 RSQVTHRMFTFPDRDELNAWQSAFSEVPILTQSAWFRPHNRDKHIDNLYLVGAGTHP 469  
 QY 472 QSGLPITYESARISAKLI 489  
 DB 470 GAGIPGVIGSAKATAGLM 487

RESULT 11  
 US-10-218-118-8  
 ; Sequence 8, Application US/10218118  
 ; Publication No. US20030148319A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bzostowicz, Patricia  
 ; APPLICANT: Rouviere, Pierre  
 ; APPLICANT: Picataggio, Stephen  
 ; APPLICANT: Cheng, Qiong  
 ; TITLE OF INVENTION: Genes Encoding Carotenoid Compounds  
 ; FILE REFERENCE: CL1876 US NA  
 ; CURRENT APPLICATION NUMBER: US/10/218,118  
 ; CURRENT FILING DATE: 2002-08-13  
 ; PRIOR APPLICATION NUMBER: 60/312,646  
 ; PRIOR FILING DATE: 2001-08-15  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 8  
 ; LENGTH: 492  
 ; TYPE: PRT  
 ; ORGANISM: Pantoea stewartii  
 US-10-218-118-8

Query Match 25.9%; Score 703.5; DB 14; Length 492;  
 Best Local Similarity 31.7%; Pred. No. 3.1e-63;  
 Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

QY 5 KHIIVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRNRPINMGFTFTDGTFFLLMKG 64  
 DB 2 KPTTVIGAGGGLALAIRLQAAGIPVLLLEQDRKPGKRAYVYQGGFTFDAGPTVITDPS 61  
 QY 65 VLDEMFCERRSBDYLFPLSPMYRLLYDDRDIFVY-SDENNRRAELQRFVDEGTDGY 123

DB 62 ALIELFALAGLQKDYVELLPVTFYRLCWESGKVFNYDNDQALEAQIOQFNPRDVAGY 121  
 QY 124 EQFMQERKRNFALPCITRDYSSLK-----SFLSL-DLI:ALPWL:---FPKSVFNNLQ 175  
 DB 122 RAFLDYSRAVNEG------LKLGTVPFLSKDMLRAAPQAKLQAWRSVYSKVG 172  
 QY 176 YFNQKMLAFQFQSKYLGMSWPCPALFTMLPYLEHEYYIYHVKGLNRIAAQAQVIA 235  
 DB 173 YIEDEHLAQAFSFLSLVGGNPPATSSIIYTLIHALEREWVMPRGCTGALVNGMKLFQ 232  
 QY 236 ENGGEIHLNSIESLIIENGAAGKVKLGHAELRGDEVIINADPAHAMTHLVK--PGVLK 293  
 DB 233 DLGGEVNLARVSHMETVGDKIQAQVLEDGRFRFTCAVSNADVVHYRDLLSQHPAAK 292  
 QY 294 KYTPENLQREYSCSTFMYLGLDKIYD-LPHHTIVFAKYTTNIRNIPFNKTLTDDFSP 352  
 DB 293 Q--AKKLQSKMSNSLFVLYFGLNHHDDQLAHTVCFGPYRELIIHFHHDGLAEDFSL 350  
 QY 353 YVQNASASDDSLAPAGKSAVYLVMPN-NDSGLDQAHQNVREOVLDTLGARLGLSDI 411  
 DB 351 YLHAPCVTDPSLAPBGCGSYVYLVAPVPHLGTANLDMAVEGRLDRIFDYLEQHY-MPGL 409  
 QY 412 RAHIECEKIIPTQWETDEHYKATFSLSHKFSQMLYWRPHNRFELANCYLVGCGTHP 471  
 DB 410 RSQVTHRMFTFPDRDELNAWQSAFSEVPILTQSAWFRPHNRDKHIDNLYLVGAGTHP 469  
 QY 472 QSGLPITYESARISAKLI 489  
 DB 470 GAGIPGVIGSAKATAGLM 487

RESULT 12  
 US-10-358-917-12  
 ; Sequence 12, Application US/10358917  
 ; Publication No. US20030182687A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cheng, Qiong  
 ; APPLICANT: No. US20030182687A1ton, Kelley C.  
 ; APPLICANT: Tao, Luan  
 ; TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOID COMPOUNDS  
 ; FILE REFERENCE: CL1929 US NA  
 ; CURRENT APPLICATION NUMBER: US/10/358,917  
 ; CURRENT FILING DATE: 2003-02-05  
 ; PRIOR APPLICATION NUMBER: 60/355,939  
 ; PRIOR FILING DATE: 2002-02-11  
 ; NUMBER OF SEQ ID NOS: 28  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 12  
 ; LENGTH: 502  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus aureus  
 US-10-358-917-12

Query Match 25.9%; Score 702.5; DB 14; Length 502;  
 Best Local Similarity 31.5%; Pred. No. 4e-63;  
 Matches 159; Conservative 108; Mismatches 196; Indels 41; Gaps 12;

QY 7 IIVVGAGPGGLCAGMLLSQRFKVSIFDKHABIGGRNRPINMGFTFTDGTFFLLMKGVL 66  
 DB 3 IAVIGAGVTGLAAARIASQGEVTFEKNNVGRMNO:KXGFTFDGPTIWMEDVY 62  
 QY 67 DEMFELCERRSBDYLFPLSPMYRLLYDDRD-IFVYSD:ENMRRAELQRFVDEGTDGYEQ 125  
 DB 63 KDVFACGKNVEDYIELQLRYVDVYFDHDDRTVPTD:AELOQMLESIPGSGTHGFWMS 122  
 QY 126 FNEQERKRNFALPCITR-----DYSLKSL:LDLIKALPWLAPFKSVFNUL 173  
 DB 123 FLTDVYKYE-----IARRYFLERTYRKPSDFYNTSLV:SAKUTL:-----NHA 167  
 QY 174 GQ-----YFNQERKRNFALPCITR-----DYSLKSL:LDLIKALPWLAPFKSVFNUL 229  
 DB 168 DQLEIHYIDNEKIQLAFQTLIYIGIDPKRGPVLSIIP:LEMMFGVHFIKGMYGMAQG 227



429 DEHYKATGATSLSHKTSOMLYWRPHNRPRELANCYLVGGTHPGSGLTPIYESARISAKL 488  
476 TLDSYLGNAFGPEPRLVQSAPFFPHNRSEDLNHNFLVGVGAGQPGAGTSPVMSAKTARL 535  
489 ISQ 491  
536 IAE 538

RESULT 15  
US-09-547-267-5  
Sequence 5, Application US/09547267  
Patent No. US20020147371A1  
GENERAL INFORMATION:  
APPLICANT: Hohmann, Hans-Peter  
APPLICANT: Pasamontes, Luis  
APPLICANT: Tessier, Michel  
APPLICANT: van Loon, Adolphus  
TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hoffmann-La Roche Inc.  
STREET: 340 Kingsland Street  
CITY: Nutley  
STATE: NJ  
COUNTRY: USA  
ZIP: 07110

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/547,267  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/660,645  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Pokras, Bruce A.  
REGISTRATION NUMBER: 32,748  
REFERENCE/DOCKET NUMBER: RAN 6002/170  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 235-5801  
TELEFAX: (201) 235-2363  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 494 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-547-267-5  
Query Match 23.6%; Score 640.5; DB 9; Length 494;  
Best Local Similarity 30.9%; Pred. No. 9.7e-57;  
Matches 150; Conservative 98; Mismatches 231; Indels 7; Gaps 6;  
2Y 8 IIVGAGPGGLCAGMLLSQGFVKSIIDKHAIEIGGRNPINWNGFTDTGPTLLMKGVLD 67  
DB 5 IIVGAGFGGLALNLQAGATTVIIVARDKPGGRAYVWDDQGHVFDAGTVVTDPSLR 64  
2Y 68 EMFELCERRSEDTLEFLPLSPMYRLYYDDRIIFYV-SDRENMRALQRFDEGTDGYEQF 126  
DB 65 ELWALSGQPMERDVTLLFVSPFYRLTWADGRSPFYVNDDELIRQVASFPNADVDGYRRF 124  
2Y 127 MEQERKRNFALYPCIT-TRDYSLSKSLFSLDLIKALPWLAPKSVFNMLGYFNQEXMLA 185  
DB 125 HDYAEVYREGYKLGIKGTTPF--LKLQGLMLNAPALMELQAYRSVHSMVAFIQDPHLRQA 182  
2Y 186 FCFQSKYLGMSPECPALFTMLPYLEHEYIYHVKGGLNRIIAAQAQVIAENGGEIHLNS 245

DB 183 FSHFTLLVGNPFSTSSIIYALTHALERRGGWPFAGGTTN LVAGMVALFERLGGTILLNA 242  
QY 246 EIESLIIEENGAAKGVKLQHGCAELRGDEVIINADPAHAMTI LVKPGVLKXYTPENLKOREY 305  
DB 243 RVTRIDTEGDRATGVTLIDGRQLRADTVASNGSDVMHSYRI LLGHTTRRGRTKAAILNRQRM 302  
QY 306 SCSTFMLYLGLDK-IYDLPHHTIVPAKDYTTNIRNIFDNH LTTDDPSFYVQNASASDDSL 364  
DB 303 SMSLFLVLFGLSKRPNLAHSHVIFGRYKGLVNEIFNGERLPDDFSMYLHSPCVTDPSL 362  
QY 365 APAGKSALYVLVPMPN-NDSGLDWQAFCONVREQVLTLCARLGLSDIRAHIECEKIITP 423  
DB 363 APEGCMSTHYVLA PVPHGLRADVDWEAEAPGYAERIPEEL ERRALPOLAKHLTVSRIFSP 421  
QY 424 QTWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEELANC LVCGGTHPGSGLTPIYESAR 483  
DB 422 ADFSTELSAHFSGSAFSVEPILATQSAWFRPHNRDRALPNF IVGAGTHPGAGIPGVVGSAR 481  
QY 484 ISAKLI 489  
DB 482 ATAQVM 487

Search completed: February 29, 2004, 15:27:50  
Job time : 37.9275 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

on: February 29, 2004, 14:35:44 ; Search time 15.5283 Seconds  
(without alignments)  
1698.885 Million cell updates/sec

file: US-09-941-947A-22

effect score: 2711

sequence: 1 MANTKHIIIVGAGPGLCAG.....KRVSRFKDIASHAWLKAKA 511

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Issued Patents AA:\*

1: /cgn2\_6/prodata/2/iaa/5A-COMB.pep:\*

2: /cgn2\_6/prodata/2/iaa/5B-COMB.pep:\*

3: /cgn2\_6/prodata/2/iaa/6A-COMB.pep:\*

4: /cgn2\_6/prodata/2/iaa/6B-COMB.pep:\*

5: /cgn2\_6/prodata/2/iaa/PCTus-COMB.pep:\*

6: /cgn2\_6/prodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	2711	100.0	511	4	US-09-934-903-16
2	700.5	25.8	492	1	US-07-783-705A-4
3	673	24.8	582	4	US-09-091-725-17
4	640.5	23.6	494	3	US-08-560-645A-5
5	640.5	23.6	494	3	US-09-298-718-5
6	640.5	23.6	494	3	US-09-546-969-5
7	640.5	23.6	494	3	US-08-980-832-4
8	640.5	23.6	494	4	US-09-547-267-5
9	640.5	23.6	494	4	US-09-920-923B-4
10	614	22.6	489	1	US-08-095-726-8
11	614	22.6	489	1	US-08-096-043-8
12	614	22.6	489	1	US-08-096-623A-8
13	611	22.5	489	1	US-08-095-726-10
14	611	22.5	489	1	US-08-096-043-10
15	611	22.5	489	1	US-08-096-623A-10
16	607.5	22.4	497	4	US-09-934-903-18
17	237.5	8.8	290	4	US-08-936-165A-462
18	206	7.6	610	4	US-09-443-184-55
19	140.5	5.2	524	3	US-09-230-388-1
20	140.5	5.2	524	4	US-09-912-176-1
21	126.5	4.7	588	4	US-09-690-942-13
22	125.5	4.6	489	4	US-09-986-536-2
23	121	4.5	354	4	US-09-198-452A-958
24	119.5	4.4	475	4	US-09-489-039A-13710
25	119.5	4.4	582	1	US-08-261-086-4
26	119	4.4	582	4	US-08-261-086-6
27	118.5	4.4	568	4	US-09-690-942-6

ALIGNMENTS

RESULT 1

US-09-934-903-16

; Sequence 16, Application US/09934903

; Patent No. 6660507

; GENERAL INFORMATION:

; APPLICANT: Koffas, J. Martin

; APPLICANT: Odum, J. Martin

; APPLICANT: Schenzle, Andreas J.

; APPLICANT: No. 6660507ton, Kelley C.

; APPLICANT: Tomb, Jean-Francois

; APPLICANT: Rouviere, Pierre

; APPLICANT: Picataggio, Stephen

; APPLICANT: Cheng, Giong

; TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production

; FILE REFERENCE: C11646 US NA

; CURRENT APPLICATION NUMBER: US/09/934,903

; CURRENT FILING DATE: 2001-08-22

; PRIOR APPLICATION NUMBER: 60/229,907

; PRIOR FILING DATE: September 1, 2001

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 16

; LENGTH: 511

; TYPE: PRT

; ORGANISM: Methylobionas 16a

; FEATURE:

; OTHER INFORMATION: Amino acid sequences encoded by ORF8

US-09-934-903-16

Query Match 100.0%; Score 2711; DB 4; Length 511;

Best Local Similarity 100.0%; Pred. No. 9,1e-266;

Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MANTKHIIIVGAGPGLCAGMLLSORGFVSVFDFKHASIC.RNRPINMGFTFDGPTFL 60

Db 1 MANTKHIIIVGAGPGLCAGMLLSORGFVSVFDFKHASIC.RNRPINMGFTFDGPTFL 60

Qy 61 LMKGVLDMEFELCERSESDYLEPLSPMYRLLYDDRDIF YSDRENMAELQVDEGT 120

Db 61 LMKGVLDMEFELCERSESDYLEPLSPMYRLLYDDRDIF YSDRENMAELQVDEGT 120

Qy 121 DGVEQFMEQERKEFNALYPCITRDYSSLSKLSFLSGLIKALPMLAPKPSVFNNGQVFNQE 180

Db 121 DGVEQFMEQERKEFNALYPCITRDYSSLSKLSFLSGLIKALPMLAPKPSVFNNGQVFNQE 180

Qy 181 KRLAFCFQSKYLGMSFPCALFTMLPYLEHYGYHYVK ELNR1AAAMAQVIAENGCE 240

Db 181 KRLAFCFQSKYLGMSFPCALFTMLPYLEHYGYHYVK ELNR1AAAMAQVIAENGCE 240

Qy 241 IHLNSEIISLIENGAAKGVKQHGAEKRGDEVIINADFA.MNTHLVKPGVLKXKYPENL 300

Db 241 IHLNSEIISLIENGAAKGVKQHGAEKRGDEVIINADFA.MNTHLVKPGVLKXKYPENL 300



Db 241 IHLNSELISLIENGAAKGVKLGABLRGDEVIINADFAAMTHLVKPGVLKXITPENL 300  
QY 301 KOREYSCSTFMYLGLDKYIDLPHETIVFAKYTTNIRNIFDNKTLTDDFSFYVQNASAS 360  
Db 301 KOREYSCSTFMYLGLDKYIDLPHETIVFAKYTTNIRNIFDNKTLTDDFSFYVQNASAS 360  
QY 361 DDLAPAGKSAIYLVLPNPNDSGLDWAHCONVREOVLDTLGLARGLSDIRAHIECEKI 420  
Db 361 DDLAPAGKSAIYLVLPNPNDSGLDWAHCONVREOVLDTLGLARGLSDIRAHIECEKI 420  
QY 421 ITPQWETDEHYKATGATSLSHKPSQMLYWRPHNRFELANCYLVGCGTHPGSGLPITYE 480  
Db 421 ITPQWETDEHYKATGATSLSHKPSQMLYWRPHNRFELANCYLVGCGTHPGSGLPITYE 480  
QY 481 SARISAKLISQKRVRFKDIASHAWLKAKA 511  
Db 481 SARISAKLISQKRVRFKDIASHAWLKAKA 511

RESULT 2  
US-07-783-705A-4  
; Sequence 4, Application US/07783705A  
; Patent No. 5429939  
; GENERAL INFORMATION:  
; APPLICANT: Misawa, No. 5429939iuhiko  
; APPLICANT: Kobayashi, Kazuo  
; APPLICANT: Nakamura, Katsumi  
; APPLICANT: Yamano, Shigeoyuki  
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE  
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ladass & Parry  
; STREET: 26 West 61 Street  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10023  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720Kb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: N/A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/783.705A  
; FILING DATE: 19911023  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 1-103078  
; FILING DATE: 21-APR-1989  
; APPLICATION NUMBER: JP 2-53225  
; FILING DATE: 05-MAR-1990  
; APPLICATION NUMBER: US 07/519,011  
; FILING DATE: 19-APR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Schwadron, Janet I.  
; REGISTRATION NUMBER: 33,778  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-708-1935  
; TELEFAX: 212-246-5959  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 492 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-783-705A-4

Query Match 25.8%; Score 700.5; DB 1; Length 492;  
Best Local Similarity 31.5%; Pred. No. 4.4e-62;  
Matches 157; Conservative 103; Mismatches 213; Indels 25; Gaps 10;

QY 5 KHLIIVGAGPGGLCAGMLLSQSGFKVYSIFDKHAEITGGRN PINMNGTFTDGPTELLMKG 64  
Db 2 KPTTVIGAGFGALAIETLQAAGIPVLLEQDPGGRA VYEDQGFTEAGPVTIDPS 61  
QY 65 VLDEMFELCERSEDIYFLPLSPMYRLLYDDRIDFVY- DRENNRAELQRFVDEGTDGY 123  
Db 62 AEBELFALAGKQLKEYVELLPVTPPYRLCWESGKVFNVDI DQTRLEAQIQCFNPRDVEGY 121  
QY 124 EQFMEQERKRFNVALYPCITRDYSSLK-----SFLSL-DLI-ALPWL- ---FPKSVFNILGQ 175  
Db 122 ROFLDYSRAVFKEGY-----LKLGTVPFLSFRDMLAAPQLAKIQAWSVYSKVAS 172  
QY 176 YFNQMKRLAFQFQSKYLGMSFPCPALFTMLPYLEHEVYIYHVKGGLNRIAAAMAQVIA 235  
Db 173 YTEDHLLQA-SFPHSLLVGNGPPATSSIYTLIHALEREW-WFPRGGTGALVQGMKLFQ 232  
QY 236 ENGSIHLNSELISLIENGAAKGVKLGABLRGDEVI NADFAAMTHLVK--PGVLK 293  
Db 233 DLGGEVILNARVSGHMETTGNKIEAIVHLEDGGRFLTQAVA NADVHTYRDLISQHPAAV 292  
QY 294 KYTPENLKOREYSCSTFMYLGLDKYID-LPHHTIVFAK YTTNIRNIFDNKTLTDDPSF 352  
Db 293 Q--SNKLOTKMSNSLFLVLYFGLNHHHDQLAHTVCFGPYRELIDELFNEHDGLAEDPSL 350  
QY 353 YVQNASDSDSLAPAGKSAIYLVLPNPNDSGLDWAHCONVREOVLDTLGLARGLSDI 411  
Db 351 YLHAPCVTDSLSLAPGCGSYVVLAPVPHLGTANLDWTVE PKLRDRIPAYLEQHY-MPGL 409  
QY 412 RAHIECEKIITPQWETDEHYKATGATSLSHKPSQMLYWRPHNFEELANCYLVGCGTHP 471  
Db 410 RSLQVTHRMFTFFPFRDQNLNAYHGSASFVEPLVTSQSAF-PHNDRKTIWNLVYVGAGTHP 469  
QY 472 GSGLPITYESARISAKLI 489  
Db 470 GAGIPGVIGSAKATAGLM 487  
RESULT 3  
US-09-091-725-17  
; Sequence 17, Application US/09091725  
; Patent No. 6329141  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Improved methods for transforming Phaffia  
; TITLE OF INVENTION: and recombinant DNA for use therein  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster llp  
; STREET: 2000 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: United States of America  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.2 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/091.725  
; FILING DATE: 23-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 95203620.0  
; FILING DATE: 22-DEC-1995  
; APPLICATION NUMBER: EP 96200943.7  
; FILING DATE: 11-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: E. Victor Donahue  
; REGISTRATION NUMBER: 35,492  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 582 amino acids



```

STREET: 340 Kingsland Street
CITY: Nutley
STATE: NJ
COUNTRY: USA
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/298,718
FILING DATE:
CLASSIFICATION:
Prior Application Data:
APPLICATION NUMBER: 08/660,645
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Pokras, Bruce A.
REGISTRATION NUMBER: 32,748
REFERENCE/DOCKET NUMBER: RAN 6002/170
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-5801
TELEFAX: (201) 235-2363
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 494 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-298-718-5

Query Match      23.6%; Score 640.5; DB 3; Length 494;
Best Local Similarity 30.9%; Pred. No. 5.3e-56;
Matches 150; Conservative 98; Mismatches 231; Indels 7; Gaps 6;

QY   8 IIVGAGPGGLCAGMLLSQRGEKVSIFDKHAEIGENRRPINMGFTDGTGFLLMKGVLD 67
DB   5 IIVGAGGGALAIRLQSAGIATTVIARDKPGGRAYVNDQGHVFAGPTVPDPSSLR 64

QY   68 EMPFCERRSEDYLEFLPLSPMYRLLYDDRDIFVY-SDRNNRAELQRFDEGTDGYEQF 126
DB   65 ELWALSQPWRDVTLTPVSPFYRLTWADGRSEFYVNDDELIRQVASFNPAVDGYRRF 124

QY   127 MEQEKKEFNALPCITRDYSSLSKSFSLDLIKALPWLAKPSVFNLCQYFNQOEKRLA 185
DB   125 HDVAEYVEYGVLKGTTFF-LKLQQLNNAAPALMRLOYSVHSVMARFIQDPHLROA 182

QY   186 FCFQSKYLKMSPECPALFTMLPYLEHYGIHVKGGLNIAAMAQVIAENGGEIHNS 245
DB   183 FSPHTLVGNPFSTSIYALIHALERRGVWFAGKGTNLVAGKVALLPERLGTTLLNA 242

QY   246 EIESLIENGAAGVKLOHGAEIRGDEVIIINADFAHAMTHLVKPGVKKYTPENLKOREY 305
DB   243 RVTRIDTEGDRATGVTLLDGRLRADTVASNGDVMSYELLGHTTTRRGRTKAAILNRQW 302

QY   306 SCSTFMILYGLDK-IYDLPHTITVFAKDVTYTNIRIPDN-TLTDDFSFYQNASASDSSL 364
DB   303 SNLSFLVHLGSKRPENLAHSVIEGPYKGVINEFNGPLPDPSMYLHSCVTDPSL 362

QY   365 APAGKALIVLPMN-NDSGLDWAHCNQVREQLVDTLGLARGESDIRAHICEKIITP 423
DB   363 APEGMTSHVLAAPVPHLRADVMEAEAFGYAEIRFEEL-ERRAIPLRLKHLTVSRIFSP 421

QY   424 QWEVTEHYKGAFTSLSHKTSOMLYNPHPNRFBELANCYLVGGTHPGSGGLPTIVESAR 483
DB   422 ADPSTELSAAHGSASFVEPIILTQSANFPNHRDRAIPNFVIVGHTHPGAGICPVVGSAX 481

QY   484 ISAKLI 489
DB   482 ATAQVM 487

```

Db 303 SMSLFLVHLGSLKRPENLAHSVIFGPRYKGLVNEIFNGPRLPDDFSKYLHSPCVTDPSL 362  
Qy 365 APAGKSALYVLVPMN-NDSGLDQAHCONVREQVLDLTGLARGLSDIRAHIECEKIITP 423  
Db 363 APEGMSHYVLAPVPHLGRADVDEARAPGYAEIRFEEL-ERRAIPDLRKLITVSRIFSP 421  
Qy 424 QTWETDEHYVKGATFSLSHKESQMLYWRPHNRPPELANCYLVGGTHPGSGGLTIYESAR 483  
Db 422 ADFSTELSAHGSAPFVEPILTSQAWFRPHNRDRAIFNFVIVGAGTHPGAGIPGVVGSAAK 481  
Qy 484 ISAKLI 489  
Db 482 ATAQVM 487

RESULT 7  
JS-08-980-832-4  
; Sequence 4, Application US/08980832B  
; Patent No. 6291204  
; GENERAL INFORMATION:  
; APPLICANT: Pasamontes, Luis  
; APPLICANT: Tsygankov, Yuri  
; TITLE OF INVENTION: Improved Fermentative Carotenoid Production  
; FILE REFERENCE: Improved Fermentative Carotenoid  
; CURRENT APPLICATION NUMBER: US/08/980.832B  
; CURRENT FILING DATE: 1997-12-01  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 4  
LENGTH: 494  
TYPE: PRT  
ORGANISM: Flavobacterium sp. R1534  
JS-08-980-832-4

Query Match 23.6%; Score 640.5; DB 3; Length 494;  
Best Local Similarity 30.9%; Pred. No. 5.3e-56;  
Matches 150; Conservative 98; Mismatches 231; Indels 7; Gaps 6;

Yy 8 IIVGAGPGGLCAGMLLSQRFKYSIFDKHAEIGGRNRPINMGFTDTGPTFLMKGVLD 67  
Db 5 IIVGAGPGGLALAIRLQSGAGIATTVIARDKPGRAYVNDQGHVFDAGFTVYVDPDSLR 64  
Yy 68 EMPFELCRRSEDYLEFLPLSPMYRLLYDDRDIFVY-SDRENMPAEIQRVFDGTYEQF 126  
Db 65 ELWALSQPMERDVTLPLVSPFFRYLTWADGRSPFYVNDDELIRQVASFNPADVDGYRRF 124  
Yy 127 MEQERKRFNALYPCI-TRDYSSLSKSFSLDLIKALPWLAFPKSVFNNLQGYFNOEKWRLA 185  
Db 125 HDYAEVYRGYKLGITTPF--LKQQLMNAAPALMRLQAYRSVHSMVAFIQDPHLROA 182  
Yy 186 FCFSQKYLGMSPWECPLFTMLPYLHEHYGIYHVGKGLNRIAAQAQVIAENGETHLS 245  
Db 183 FSPHTLVGNGPSTSIYALIHALLERRGGVWPAKGGTQVLVAGMVALPERLGGTLLNA 242  
Yy 246 EIESLIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLKYYTPENLKQRY 305  
Db 243 RVTRIDTEGDRATGVTLDRQRLRADTVASNGDWMHSYRDLGHTRRGRTKAAILNRQW 302  
Yy 306 SCSTFMVLYGLDK-IYDLPHHTIVPAKDYTTNIRNIFDNKLTDDPSFYVQNASDDSL 364  
Db 303 SMSLFLVHLGSLKRPENLAHSVIFGPRYKGLVNEIFNGPRLPDDFSKYLHSPCVTDPSL 362  
Yy 365 APAGKSALYVLVPMN-NDSGLDQAHCONVREQVLDLTGLARGLSDIRAHIECEKIITP 423  
Db 363 APEGMSHYVLAPVPHLGRADVDEARAPGYAEIRFEEL-ERRAIPDLRKLITVSRIFSP 421  
Yy 424 QTWETDEHYVKGATFSLSHKESQMLYWRPHNRPPELANCYLVGGTHPGSGGLTIYESAR 483  
Db 422 ADFSTELSAHGSAPFVEPILTSQAWFRPHNRDRAIFNFVIVGAGTHPGAGIPGVVGSAAK 481  
Yy 484 ISAKLI 489  
Db 482 ATAQVM 487

## RESULT 8

US-09-547-267-5  
; Sequence 5, Application US/09547267  
; Patent No. 6613543  
; GENERAL INFORMATION:  
; APPLICANT: Hohmann, Hans-Peter  
; APPLICANT: Pasamontes, Luis  
; APPLICANT: Tessier, Michel  
; APPLICANT: van Loon, Adolphus  
; TITLE OF INVENTION: FERMENTATIVE CAROTENOID PRODUCTION  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hoffmann-La Roche Inc.  
; STREET: 340 Kingsland Street  
; CITY: Nutley  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/547,267  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/660,645  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pokras, Bruce A.  
; REGISTRATION NUMBER: 32,748  
; REFERENCE/POCKET NUMBER: RAN 6002/170  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (201) 235-5801  
; TELEFAX: (201) 235-2363  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 494 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-547-267-5

Query Match 23.6%; Score 640.5; DB 4; Length 494;  
Best Local Similarity 30.9%; Pred. No. 5.3e-56;  
Matches 150; Conservative 98; Mismatches 231; Indels 7; Gaps 6;

Qy 8 IIVGAGPGGLCAGMLLSQRFKYSIFDKHAEIGGRNRPINMGFTDTGPTFLMKGVLD 67  
Db 5 IIVGAGPGGLALAIRLQSGAGIATTVIARDKPGRAYVNDQGHVFDAGFTVYVDPDSLR 64  
Qy 68 EMPFELCRRSEDYLEFLPLSPMYRLLYDDRDIFVY-SDRENMPAEIQRVFDGTYEQF 126  
Db 65 ELWALSQPMERDVTLPLVSPFFRYLTWADGRSPFYVNDDELIRQVASFNPADVDGYRRF 124  
Qy 127 MEQERKRFNALYPCI-TRDYSSLSKSFSLDLIKALPWLAFPKSVFNNLQGYFNOEKWRLA 185  
Db 125 HDYAEVYRGYKLGITTPF--LKQQLMNAAPALMRLQAYRSVHSMVAFIQDPHLROA 182  
Qy 186 FCFSQKYLGMSPWECPLFTMLPYLHEHYGIYHVGKGLNRIAAQAQVIAENGETHLS 245  
Db 183 FSPHTLVGNGPSTSIYALIHALLERRGGVWPAKGGTQVLVAGMVALPERLGGTLLNA 242  
Qy 246 EIESLIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLKYYTPENLKQRY 305  
Db 243 RVTRIDTEGDRATGVTLDRQRLRADTVASNGDWMHSYRDLGHTRRGRTKAAILNRQW 302  
Qy 306 SCSTFMVLYGLDK-IYDLPHHTIVPAKDYTTNIRNIFDNKLTDDPSFYVQNASDDSL 364

Db 303 SMLFVLFHFGSLKRPENLAHRSVIFGPKYGLVNEIFNGFRLPDDFSMYLHSPCVTDPSL 362  
QY 365 APAGKSAIYVLVPMFN-NDUGLDWQACQVQVEQVLDTLGRLGLSDIRAHIEKEKIITP 423  
Db 363 APGGMSTRYVLAPVPHLGRADVWEABAPGYAERIFEL-ERRAIPDLRKLTVSRIFSP 421  
QY 424 QTWETDEHYKATGATSLSKFSQMLYWRPHNPFEBLANCYLVGGCTHFGSGLTIVYESAR 483  
Db 422 ADPSTELSAHGSFAFSVEPILTQSAWFRPHNRDRAIPNFYIVGAGTHFGAGIPGVWGSAX 481  
QY 484 ISAKLI 489  
Db 482 ATAQVM 487

## RESULT 9

US-09-920-923B-4  
; Sequence 4, Application US/09920923B  
; Patent No. 6677134  
; GENERAL INFORMATION:  
; APPLICANT: Pasamontes, Luis  
; APPLICANT: Tsygankov, Yuri  
; TITLE OF INVENTION: Fermentative Carotenoid Production.  
; FILE REFERENCE: 15464 US (C38435/125944)  
; CURRENT APPLICATION NUMBER: US/09/920,923B  
; CURRENT FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 08/980,832  
; PRIOR FILING DATE: 1997-12-01  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 4  
; LENGTH: 494  
; TYPE: PRT  
; ORGANISM: Flavobacterium sp. R1534  
US-09-920-923B-4

Query Match 23.6%; Score 640.5; DB 4; Length 494;  
Best Local Similarity 30.9%; Pred. No. 5.3e-56;  
Matches 150; Conservative 98; Mismatches 231; Indels 7; Gaps 6;  
QY 8 IVGAGPGGLCAGMLLSQSGFKVSIPTDKHAEIGGRNRPINNGFTDGTGPTFLMKGVLD 67  
Db 5 IVGAGFGGLALAIRLQAGIATTIVEARDKPGGRAYVYNDQGHVFDAGPTVVDPSLR 64  
QY 68 EMFELCERSEDEFLPLSPMYRLLYDDRDIFVY-SDRENRAELQRFVDEGTDGYEQF 126  
Db 65 ELWALSQCPMERDVTLLPVSPYRLTWADGRSFYVNDDELIRQVASFNPADVDGTYRF 124  
QY 127 MEQERKRNFALYPCI-TRDYSSLSKSLFLDLIKALPWLAFPKSVFNILQGYFNOERKRLA 185  
Db 125 HOYAEVVRREGYLKLGTTPTF--LKLQMLNAPALMRLOQAVRSVHSMVAREIQDPHLRQA 182  
QY 186 FCFQSKYLGMSWECFALFTMLPYLEHYGYHYKGLNRIRAAAMAQVIAENGGEIHLNS 245  
Db 183 FGFHLLVGNPNFTSSYIALHALERGGVWFAGKGNQLVAGVVALFELGCTGLINA 242  
QY 246 EESIIENGAAKGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLYKXYPENLKOREY 305  
Db 243 RVTRIDTEGRATGVTLLDGRQLRADTVASGCDVMSYRDLGHTGRGRTKAALNRQRW 302  
QY 306 SCSFMVLYGLDK-ITYDLPHITVFAKYDTTNRNIFUNKLTLDGDFPYONASASDLSL 364  
Db 303 SMLFVLFHFGSLKRPENLAHRSVIFGPKYGLVNEIFNGFRLPDDFSMYLHSPCVTDPSL 362  
QY 365 APAGKSAIYVLVPMFN-NDUGLDWQACQVQVEQVLDTLGRLGLSDIRAHIEKEKIITP 423  
Db 363 APEGHSTHYVLAPVPHLGRADVWEABAPGYAERIFEL-ERRAIPDLRKLTVSRIFSP 421  
QY 424 QTWETDEHYKATGATSLSKFSQMLYWRPHNPFEBLANCYLVGGCTHFGSGLTIVYESAR 483  
Db 422 ADPSTELSAHGSFAFSVEPILTQSAWFRPHNRDRAIPNFYIVGAGTHFGAGIPGVWGSAX 481

QY 484 ISAKLI 489  
Db 482 ATAQVM 487  
RESULT 10  
US-08-095-726-8  
; Sequence 8, Application US/08095726  
; Patent No. 5530188  
; GENERAL INFORMATION:  
; APPLICANT: Ausich, Rodney L  
; APPLICANT: Brinkhaus, Friedhelm L  
; APPLICANT: Mukharji, Indrani  
; APPLICANT: Proffitt, John H  
; APPLICANT: Yarger, James G  
; APPLICANT: Yen, Huel-Che B  
; TITLE OF INVENTION: Beta-Carotene Biosynthesis in  
; TITLE OF INVENTION: Genetically Engineered Hosts  
; NUMBER OF SEQUENCES: 79  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
; STREET: 200 E Randolph St  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60680-0703  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent in Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/095,726  
; FILING DATE: 21-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/785,566  
; FILING DATE: 30-OCT-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Galloway, No. 5530188val B  
; TELEPHONE: 3128567180  
; TELEFAX: 3128564972  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 489 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-095-726-8  
Query Match 22.6%; Score 614; DB 1; Length 489;  
Best Local Similarity 30.3%; Pred No. 2.5e-53;  
Matches 152; Conservative 97; Mismatches 218; Indels 34; Gaps 12;  
QY 5 KHIIVGAGPGGLCAGMLLSQSGFKVSIPTDKHAEIGGRNRPINNGFTDGTGPTFLMKG 64  
Db 2 KKTWVIGAGFGGLALAIRLQAGIPTVLLEQDKPGGRAYVYNDQGHVFDAGPTVVDPT 61  
QY 65 VLDEMFCERSEDEFLPLSPMYRLLYDDRDIFVY-SDRENRAELQRFVDEGTDGY 123  
Db 62 ALERLFTTAGRMEDYVRLPVKPYRLCWSGKTLQVADSFLEAQITQFNPRDVEGY 121  
QY 124 EGFMEQERKRNFALYPCI-TRDYSSLSKSLFLDLIKALPWLAFPKSVFNILQGYFNO 173  
Db 122 RRLFAYSQAVFQ-----EGYLRGVSVPFLSRFDMRLRQAPQLKLQAW-----QSVYQSV 170  
QY 174 GYFQSKYLGMSWECFALFTMLPYLEHYGYHYKGLNRIRAAAMAQVIAENGGEIHLNS 233  
Db 171 SFPIEDEHLRQAFSPHLLVGNPNFTSSYIALHALERGGVWFAGKGNQLVAGVVALFEL 230  
QY 234 IAEENGGEIHLNSEIIESLIENGAAKGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLYK 293

b 231 FTDLGGEIELNARVEELVADNRVSOVRLADGRI FDTDAVSNADVNTYKLLGTIPVG 290  
y 294 KYTPENLKOREYSCSTFMYLGLDKIY-DLPHHTIVFAKDYTTNIRNIFDNKTLTDDPSF 352  
b 291 OKBAARLERKSNLSLVLYFGLNQPHSLAHTTCFGRPRYELDEIFTGSALADDFSL 350  
y 353 YVONASASDDSLAPACKSALYVLYVPMN-NDSGLOKQAHQNVREQVLDTLGARLGLSDI 411  
b 351 YLHSPCVTDFSLAPPCASFYVLAHPVHLGNAPLDWAQBGPKLRDRIFDYLEERY-MPGL 409  
y 412 RAHIECKIITPTWETDEHVY---KGATFSLSHKFSOMLYWRPHNRPEELANCYLVGCG 468  
b 410 RSLQVTRIFTRQ---TSHAWIALGSLFIEPPSLTQGLFAANATRH---SNLYLVAAG 463  
y 469 THPGSGLPTIYESARISAKLI 489  
b 464 THPGAGIPGVWGLAESTASLM 484

## RESULT 11

S-08-096-043-8

Sequence 8, Application US/08096043

Patent No. 5530189

## GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L  
APPLICANT: Brinkhaus, Friedhelm L  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H  
APPLICANT: Yarger, James G  
APPLICANT: Yen, Huel-Che B  
TITLE OF INVENTION: Lycopene Biosynthesis in  
TITLE OF INVENTION: Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
STREET: 200 E Randolph St  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60680-0703

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096.043  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/785,568  
FILING DATE: 30-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, No. 5530189val B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 3128567180  
TELEFAX: 3128564372

## INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:  
LENGTH: 489 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
I-08-096-043-8

## Query Match

Best Local Similarity 22.6%; Score 614; DB 1; Length 489;

Matches 152; Conservative 97; Mismatches 218; Indels 34; Gaps 12;

5 KHIIVGAPGGLCAGYLLSQRGFKVIFDKHAEIGGRNRPINNGFTPTGTFFILMKG 64

2 KKTIVGAGFGGLAIFRLQAGIPTVLLEQRDKRGGRAYVWHDDQGFDTAGTIVTIDPT 61

QY 65 VLDMFELCERSEDEYLEFLPLSPMYRLLYDDRDIFVY-SURENMRALQORVFEDEGTQY 123  
Db 62 ALBALFTLAGRMEDYVRLLPVKFPYRLCKESGKTLDYANESFELEAQITQFNPRDVEGY 121  
QY 124 EQMGEQRFRNALYPCITRDYSSLSK---FISL-DLIKALF-----WLAPPKSVFNML 173  
Db 122 RREFLAYSOAVFQ-----EGYLRGSGVFLSFRDMLRAGIQLKLQAW---QSVYQSV 170  
QY 174 GOYFNOEKVRLAFPCFOSKYLGMSPECPALFTMLPYLBEHGIYHVKGGLNRIAAQAQV 233  
Db 171 SRFIDEHLRQAFSHSLVGGNPTTSIYLIHALERENGWFPFGGTGALVNGWVKL 230  
QY 234 IANNGEIHNSIEISLIENGAAGVKLOHGAELRGDEVIINADFAHAMTHLVKPGVLK 293  
Db 231 FTDLGGEIELNARVEELVADNRVSOVRLADGRI FDTDAVSNADVNTYKLLGTIPVG 290  
QY 294 KYTPENLKOREYSCSTFMYLGLDKIY-DLPHHTIVFAKDYTTNIRNIFDNKTLTDDPSF 352  
Db 291 OKBAARLERKSNLSLVLYFGLNQPHSLAHTTCFGRPRYELDEIFTGSALADDFSL 350  
QY 353 YVONASASDDSLAPACKSALYVLYVPMN-NDSGLOKQAHQNVREQVLDTLGARLGLSDI 411  
Db 351 YLHSPCVTDFSLAPPCASFYVLAHPVHLGNAPLDWAQBGPKLRDRIFDYLEERY-MPGL 409  
QY 412 RAHIECKIITPTWETDEHVY---KGATFSLSHKFSOMLYWRPHNRPEELANCYLVGCG 468  
Db 410 RSLQVTRIFTRQ---TSHAWIALGSLFIEPPSLTQGLFAANATRH---SNLYLVAAG 463  
QY 469 THPGSGLPTIYESARISAKLI 489  
Db 464 THPGAGIPGVWGLAESTASLM 484

## RESULT 12

US-08-096-623A-8

Sequence 8, Application US/08096623A

Patent No. 5684238

## GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L  
APPLICANT: Brinkhaus, Friedhelm L  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H  
APPLICANT: Yarger, James G  
APPLICANT: Yen, Huel-Che B  
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and  
TITLE OF INVENTION: Glycosylated Zeaxanthin in Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Welsh & Katz, Ltd.  
STREET: 120 S. Riverside Plaza, 22nd Floor  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096.623A  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,061  
FILING DATE: 09-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/562,921  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/562,674  
FILING DATE: 03-AUG-1990  
PRIOR APPLICATION DATA:



467 GAGIPGVVGLAESTASLM 484

RESULT 14

IS-08-096-043-10  
Sequence 10, Application US/08096043  
Patent No. 5530189

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.  
APPLICANT: Brinkhaus, Friedhelm L.  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H.  
APPLICANT: Yarger, James G.  
APPLICANT: Yen, Hwei-Che B.  
TITLE OF INVENTION: Lycopen Biosynthesis in Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amoco Corp., Patents and Licensing Dept  
STREET: 200 E Randolph St  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60680-0703

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,043  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/785,568  
FILING DATE: 30-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Galloway, No. 5530189val B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 3128567180  
TELEFAX: 3128564972  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 489 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein

IS-08-096-043-10

Query Match 22.5%; Score 611; DB 1; Length 489;  
Best Local Similarity 30.3%; Pred. No. 5e-53;  
Matches 151; Conservative 97; Mismatches 216; Indels 34; Gaps 12;

8 IIVGAGPGLCAGMILLSQRFKVSIFDKHAEIGGRNRPINMGFTPTDGTFLAKGVLD 67  
5 VVIGAGPGLALAIRLQAAGIPTVLEQKQGRAYVWHDDQGFDTAGPTVITDPTALE 64  
68 EMPELCERSSEDLLEPLPLSPMYRLLYDDRDIFVY-SDRENRAELQRVDSGTGYEQF 126  
65 ALFTLAGRMEDVRLLPVAPFTFLCWESGKLDYANDSFELEAQITQNPDRDVGTRP 124  
127 MEQERKEFNALYPCITRDYSSLSKSL-PLSL-DLIKALP-----WLAPKPSVFNLCQY 176  
125 LAYSQAVFQ-----EGLRLGVSFPLSFRDMLRAGPQLKLQAW---QSVYQSVSRF 173  
177 FNQEKMLAFCSQKYLGMSPWPCALFTMLPYLEHEYGYHVKGLNRIIAAQAQVIAE 236  
174 IEDHLEQAFSFTSLVGGNPFPTSSYITLIHALEREWGVWPPGGTGALVNGVYKLFDT 233  
237 NGGEIHLNSESIESLIINGAAKGVKQHGAEELRGDEVIINADPAHAMTHLVKPGVLYKYT 296  
234 LGGEIELNARVEELVWADNRVSRQLADCEIFDTDAVASADWNTYKKLGLGTIPVGQKE 293

QY 297 PENLKQREYSCSTFMYLGLGLKLY-DLPHTTIVFAKDYTTNIRNIFDNKTLTDDPSFYVQ 355  
DB 294 AARLERKSMNSLFLVYFGLNQPHSQLAHTTICGPRYRELDEIFGSGALADDFSLYLH 353  
QY 356 NASASDDSLAPAGKALYLVPMNP-NDGSLDQAHQCONVREQVLDTLGARLGLSIRAH 414  
DB 354 SPCVTDPSLAPPCASFYVLAPVPHLGNAPLDWAQEGPKLDRIFDYLEERY-MPGILRSQ 412  
QY 415 IECEKIITPQWTEDSHVY---KGATFSLSHKFSOMLYNEPHNRPEELANCYLVGGGTHP 471  
DB 413 LVQRIPTRO---TSRHAWTAIIGLSLFIIEPPSLTQGLFPAANATH---SNLYLVAAGTHP 466  
QY 472 GSGLPITVESARISAKLI 489  
DB 467 GAGIPGVVGLAESTASLM 484

RESULT 15

US-08-096-623A-10  
Sequence 10, Application US/08096623A  
Patent No. 5684238

GENERAL INFORMATION:

APPLICANT: Ausich, Rodney L.  
APPLICANT: Brinkhaus, Friedhelm L.  
APPLICANT: Mukharji, Indrani  
APPLICANT: Proffitt, John H.  
APPLICANT: Yarger, James G.  
APPLICANT: Yen, Hwei-Che B.  
TITLE OF INVENTION: Biosynthesis of Zeaxanthin and Glycosylated Zeaxanthin in Genetically Engineered Hosts  
NUMBER OF SEQUENCES: 104  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Welsh & Katz, Ltd.  
STREET: 120 S. Riverside Plaza, 22nd Floor  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/096,623A  
FILING DATE: 22-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/805,061  
FILING DATE: 09-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/662,921  
FILING DATE: 28-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/562,674  
FILING DATE: 03-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/525,551  
FILING DATE: 18-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/487,613  
FILING DATE: 02-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Gamson, Edward P.  
REGISTRATION NUMBER: 29,381  
REFERENCE/DOCKET NUMBER: AMO-006.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 655-1500  
TELEFAX: (312) 655-1501  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 489 amino acids  
TYPE: amino acid



Search completed: February 29, 2004, 14:54:51  
Job time : 18.5283 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

v protein - protein search, using sw model

run on: February 29, 2004, 14:26:38 ; Search time 60.417 Seconds  
(without alignments)  
2389.754 Million cell updates/sec

file: US-09-941-947A-22

effect score: 2711  
sequence: 1 MANTKHIIIVGAGPGGLCAG.....KRVRFKDIASHAWLKAKA 511

coring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

sarched: 1586107 seqs, 282547505 residues

total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

atabase : A Geneseq 29Jan04: \*  
1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	2711	100.0	511	5	Aae22309 Methylo
2	2711	100.0	511	5	Abg61588 High grow
3	2711	100.0	511	5	Aau80332 Methylo
4	2711	100.0	511	6	Adal1434 Methylo
5	929.5	34.3	490	5	Abu49224 Listeria
6	929.5	34.3	490	6	Abu32531 Protein e
7	783.5	28.9	499	6	Aar64269 Anabaena
8	713	26.3	494	6	Abm70123 Phototrab
9	703.5	25.9	492	5	Aae22314 Pantoea s
10	703.5	25.9	492	6	Aaol16021 Pantoea s
11	703.5	25.9	492	6	Abp96688 Pantoea s
12	702.5	25.9	502	6	Abm72613 Staphyloc
13	702.5	25.9	502	6	Adal14542 Staphyloc
14	700.5	25.8	492	2	Aar07466 Polypepti
15	700.5	25.8	492	2	Aaw82257 C. utilis
16	700.5	25.8	492	2	Aaw87889 Protein e
17	700.5	25.8	492	2	Aaw99099 Erwinia u
18	700.5	25.8	492	2	Aay26333 Erwinia u
19	673	24.8	582	2	Aaw22499 Phaffia d
20	651	24.0	582	6	Abp97465 Blakeslea
21	640.5	23.6	494	2	Aaw00871 Flavobact
22	638.5	23.6	494	2	Aaw69532 Flavobact
23	624.5	23.0	526	2	Aar95697 Erythroba
24	614	22.6	489	2	Aaw01122 Phytoene
25	614	22.6	489	2	Aaw00173 Phytoene

26	614	22.6	489	2	Aaw32472	Aaw32472 Erwinia h
27	611	22.5	489	2	Aar13985	Aar13985 Phytoene
28	611	22.5	489	2	Aar13984	Aar13984 Phytoene
29	611	22.5	489	2	Aaw01123	Aaw01123 Phytoene
30	611	22.5	489	2	Aaw00174	Aaw00174 Phytoene
31	611	22.5	489	2	Aaw32475	Aaw32475 Erwinia h
32	607.5	22.4	497	5	Aae22310	Aae22310 Methylo
33	607.5	22.4	497	5	Abg61589	Abg61589 High grow
34	607.5	22.4	497	5	Aau80333	Aau80333 Methylo
35	607.5	22.4	497	6	Adal14538	Adal14538 Methylo
36	559.5	20.6	502	6	Abu43877	Abu43877 Protein e
37	502.5	18.5	497	6	Abu16031	Abu16031 Protein e
38	500.5	18.5	497	6	Adal14544	Adal14544 Staphyloc
39	500.5	18.5	500	6	Abm72616	Abm72616 Staphyloc
40	483	17.8	548	4	Aab85728	Aab85728 Enzyme in
41	482.5	17.8	544	5	Aaol15518	Aaol15518 Agromyces
42	480	17.7	548	4	Aab76640	Aab76640 Corynebac
43	480	17.7	548	4	Aab76641	Aab76641 Corynebac
44	480	17.7	548	4	Aag90440	Aag90440 C glutami
45	447.5	16.5	439	4	Aau34200	Aau34200 Staphyloc

#### ALIGNMENTS

RESULT 1  
AAE22309  
ID AAE22309 standard; protein; 511 AA.

XX AAE22309;

DT 07-AUG-2003 (revised)

DT 25-JUL-2002 (first entry)

XX Methylomonas 16a sp. diapophytoene dehydrogenase (CrtnI) enzyme.

XX Carotenoid; isopentenyl pyrophosphate; anthraxanthin; astaxanthin; diet;  
XX anti-oxidant; steroid; flavour; fragrance; electro-optic application;  
XX aquaculture; enzyme; diapophytoene dehydrogenase; CrtnI.

XX Methylomonas sp.

XX WO200218617-A2.

XX 07-MAR-2002.

XX 04-SEP-2001; 2001WO-US027420.

XX 01-SEP-2000; 2000US-0229858P.

XX 01-SEP-2000; 2000US-0229907P.

XX (DUPO ) DU PONT DE REMOURS & CO E I.

XX Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX Odum JM, Picataggio SK, Rouviere PE;

XX WPI; 2002-351711/38.

XX N-PSDB; AAD35507.

XX Producing carotenoid compounds e.g. anthraxanthin and astaxanthin, by  
XX using microorganisms having a nucleic acid molecule encoding enzymes in  
XX the carotenoid biosynthetic pathway and which metabolize single carbon  
XX substrates.

XX Claim 37; Page 127-129; 156pp; English.

XX The invention relates to a method for producing carotenoid compounds. The  
XX method comprises a transformed metabolising host cell, comprising  
XX suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule  
XX encoding an enzyme in the carotenoid biosynthetic pathway, under the  
XX control of regulatory sequences, and contacting the host cell with carbon  
XX substrate to produce a carotenoid compound. The method is useful for  
XX producing carotenoid compounds such as anthraxanthin and astaxanthin, by

using microorganism having a nucleic acid molecule encoding enzymes in the carotenoid biosynthetic pathway and which metabolize single carbon substrates. The carotenoids have potent anti-oxidant properties useful in diet, and aquaculture elements. The carotenoids are also useful as intermediates in the synthesis of steroids, flavours and fragrances and compounds for potential electro-optic applications. The present sequence is *Methylobacterium* 16a sp. diaphotocene dehydrogenase (CrtN1) enzyme used in the invention. (Updated on 07-AUG-2003 to correct OS field.)

XX Sequence 511 AA;

Query Match 100.0%; Score 2711; DB 5; Length 511;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-265;  
 Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MANTKHIIIVGAGPGGLCAGMLLSQSGFKVSIPOKHAIEIGGRNPINNMGTFTDTGPTFL 60  
 1 MANTKHIIIVGAGPGGLCAGMLLSQSGFKVSIPOKHAIEIGGRNPINNMGTFTDTGPTFL 60  
 61 LMKGVLDENFELCERSEDIYLFELPSMYRLLYDDROIIVYSRENWRAELQRFDEGT 120  
 61 LMKGVLDENFELCERSEDIYLFELPSMYRLLYDDROIIVYSRENWRAELQRFDEGT 120  
 121 DGYEQFMEQERKRFNALYPCITRDYSSLSKSFSLDLIKALPFWLAPFKSVFNNGQYFNQE 180  
 121 DGYEQFMEQERKRFNALYPCITRDYSSLSKSFSLDLIKALPFWLAPFKSVFNNGQYFNQE 180  
 181 KMLAFCFQSKYLGMSWPECPCALFTMLPYLEHYGIYHVQGLNRIIAAQAQVIAENGGE 240  
 181 KMLAFCFQSKYLGMSWPECPCALFTMLPYLEHYGIYHVQGLNRIIAAQAQVIAENGGE 240  
 241 IHLNSESIIENGAAGVKLQHGAEIRGDEVIINADFAHAMTHLVKPGVLKYYTPENL 300  
 241 IHLNSESIIENGAAGVKLQHGAEIRGDEVIINADFAHAMTHLVKPGVLKYYTPENL 300  
 301 KOREYSCSTFMYLGLDKIYDLPHTIIVFAKYDTNINIPDNKLTDDPSFYVQNASAS 360  
 301 KOREYSCSTFMYLGLDKIYDLPHTIIVFAKYDTNINIPDNKLTDDPSFYVQNASAS 360  
 361 DDSLAPAGKSAIYLVPMNPNDGLDQAHQCNVREQLDTLQARGLSDIRAHICEKI 420  
 361 DDSLAPAGKSAIYLVPMNPNDGLDQAHQCNVREQLDTLQARGLSDIRAHICEKI 420  
 421 ITPQWETDEHYVKGATSLSHKFSOMLYWPHRFRFELANCYLVGGTHPGSGLEIYE 480  
 421 ITPQWETDEHYVKGATSLSHKFSOMLYWPHRFRFELANCYLVGGTHPGSGLEIYE 480  
 481 SARISAKLISQHRVRFKDIASHAWLKZAKA 511  
 481 SARISAKLISQHRVRFKDIASHAWLKZAKA 511

# RESULT 2

ABG61588  
 ID ABG61588 standard; protein; 511 AA.

XX

AC ABG61588;

XX

DT 07-AUG-2003 (revised)

DE 27-AUG-2002 (first entry)

XX

DE High growth methanotrophic bacterial strain polypeptide #38.

XX

KW High growth methanotrophic bacterial strain; C1 carbon substrate; enzyme;

KW methane; methanol; Embden-Meyerhof carbon flux pathway; 16S rRNA;

KW pyrophosphate dependent phosphofructokinase; nitrogen-containing compound;

KW ammonia; nitrate; nitrite; nitrogen; pigment; oxygen; landfill;

KW methanol-containing environment; waste water treatment system; isoprenoid;

KW nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide.

OS *Methylobacter* sp.

XX

XX WC2002020728-A2.

PN

XX

PD 14-MAR-2002.

XX

XX 28-AUG-2001; 2001MO-US026827.

XX

XX 01-SEP-2000; 2000US-0229858P.

PR

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX

XX Koffas M, Odom JM, Schenzle A;

XX

XX WPI; 2002-454358/48.

DR

DR N-PSDB; ABK33267.

XX

PT New high growth methanotrophic bacterial strain, useful for producing single cell proteins, grows on a C1 carbon substrate, and comprises a functional gene encoding in Embden-Meyerhof carbon pathway.

PT

XX Disclosure; Page 150-152; 157pp; English.

PS

XX The invention relates to a high growth methanotrophic bacterial strain, which grows on a C1 carbon substrate e.g. methane and methanol, and comprises a functional Embden-Meyerhof carbon flux pathway comprising a gene coding a pyrophosphate dependent phosphofructokinase enzyme or a 16S rRNA. The bacterial strain is useful for the production of single cell protein and for the biotransformation of a nitrogen-containing compound, e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the production of a feed product comprising a protein, carbohydrates and a pigment and for reducing oxygen demand, for removing nitrates and nitrites in methane-containing environments such as landfills, waste water treatment systems or anywhere that methane, oxygen and nitrates are present. The bacterial strain of the invention can be used as a denitrifying agent for the conversion of nitrate or nitrite to nitrous oxide with methane or methanol as a carbon source. It is also used in the production of biomass including proteins, carbohydrates and a wide variety of pigments (particularly for isoprenoid pigments for the purpose of generating animal feeds), in production of terpenoid and carotenoid compounds, useful as pigments and as monomers in polymeric materials and in production of exopolysaccharides at high levels. Sequences ABG61551-ABG61590 represent high growth methanotrophic bacterial strain proteins of the invention. (Updated on 07-AUG-2003 to correct OS field.)

SQ Sequence 511 AA;

Query Match 100.0%; Score 2711; DB 5; Length 511;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-265;  
 Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MANTKHIIIVGAGPGGLCAGMLLSQSGFKVSIPOKHAIEIGGRNPINNMGTFTDTGPTFL 60

DB 1 MANTKHIIIVGAGPGGLCAGMLLSQSGFKVSIPOKHAIEIGGRNPINNMGTFTDTGPTFL 60

OY 61 LMKGVLDENFELCERSEDIYLFELPSMYRLLYDDROIIVYSRENWRAELQRFDEGT 120

DB 61 LMKGVLDENFELCERSEDIYLFELPSMYRLLYDDROIIVYSRENWRAELQRFDEGT 120

OY 121 DGYEQFMEQERKRFNALYPCITRDYSSLSKSFSLDLIKALPFWLAPFKSVFNNGQYFNQE 180

DB 121 DGYEQFMEQERKRFNALYPCITRDYSSLSKSFSLDLIKALPFWLAPFKSVFNNGQYFNQE 180

OY 181 KMLAFCFQSKYLGMSWPECPCALFTMLPYLEHYGIYHVQGLNRIIAAQAQVIAENGGE 240

DB 181 KMLAFCFQSKYLGMSWPECPCALFTMLPYLEHYGIYHVQGLNRIIAAQAQVIAENGGE 240

OY 241 IHLNSESIIENGAAGVKLQHGAEIRGDEVIINADFAHAMTHLVKPGVLKYYTPENL 300

DB 241 IHLNSESIIENGAAGVKLQHGAEIRGDEVIINADFAHAMTHLVKPGVLKYYTPENL 300

OY 301 KOREYSCSTFMYLGLDKIYDLPHTIIVFAKYDTNINIPDNKLTDDPSFYVQNASAS 360

DB 301 KOREYSCSTFMYLGLDKIYDLPHTIIVFAKYDTNINIPDNKLTDDPSFYVQNASAS 360

OY 361 DDSLAPAGKSAIYLVPMNPNDGLDQAHQCNVREQLDTLQARGLSDIRAHICEKI 420

||||| 361 DDLAPAGKSAALYLVLPMPNNDGLDWAHCONVREQVLDTLGARLGSLDIRAHIECEKI 420  
 ||||| 421 ITPTQWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEEELANCYLVGGTTHPGSGLPTIYE 480  
 ||||| 421 ITPTQWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEEELANCYLVGGTTHPGSGLPTIYE 480  
 ||||| 481 SARISAKLISQHRVRFKDIHAHSAWLKKAKA 511  
 ||||| 481 SARISAKLISQHRVRFKDIHAHSAWLKKAKA 511

## RESULT 3

AAU80332  
 AAU80332 standard; protein; 511 AA.

AAU80332;

15-JUL-2002 (first entry)

Methylomonas 16a ORF8 crtN1 protein sequence.

Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;  
 keratinoid; pigment; flavour; fragrance; open reading frame 8; ORF8;  
 crtN1; crtN copy1; diaphytoene dehydrogenase enzyme.

Methylomonas sp.

WO200220733-A2.

14-MAR-2002.

29-AUG-2001; 2001WO-US026852.

01-SEP-2000; 2000US-02299078.

(DUPO) DU PONT DE NEMOURS & CO E I.

Cheng Q, Koffas M, Norton KC, Odum JM, Picataggio SK;

Rouviere PE, Schenzle A, Tomb J;

WPI; 2002-383051/41.

N-PSDB; ABK50088.

Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,  
 isolated from Methylomonas 16a, useful for the production of isoprenoid  
 compounds.

Claim 4; Page 78-80; 84pp; English.

The present invention relates to a new nucleic acid molecule encoding an  
 isoprenoid biosynthetic enzyme isolated from Methylomonas 16a. The  
 invention is useful for obtaining a nucleic acid molecule encoding an  
 isoprenoid compound biosynthetic enzyme, and for the microbial production  
 of isoprenoid compounds. The molecules of the invention are also useful  
 for regulating isoprenoid biosynthesis in an organism and for producing  
 recombinant organisms for producing various isoprenoid compounds. The  
 nucleic acid is also useful for feed additive, for the production of  
 keratinoids and their derivatives, isoprenoid intermediates, and as pure  
 products useful as pigments, flavours and fragrances. The present amino  
 acid sequence represents the Methylomonas 16a open reading frame 8 (ORF8)  
 crtN1 ((crtN copy1) copy 1 of diaphytoene dehydrogenase enzyme) protein  
 of the invention, as described above

Sequence 511 AA;

Query Match 100.0%; Score 2711; DB 5; Length 511;

Best Local Similarity 100.0%; Pred. No. 6.4e-265;

Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MANTKHIIIVGAGPGGLCAGMLLSQSGFVSIFDKHAIIGGNRPINMNGFTFTGTPTFL 60

1 MANTKHIIIVGAGPGGLCAGMLLSQSGFVSIFDKHAIIGGNRPINMNGFTFTGTPTFL 60

QY 61 LMKGVLDENFELCERRSEDYLFPLSPMYRLLYDDRDIFVYSORENRAELQRFVDEGT 120  
 DB 61 LMKGVLDENFELCERRSEDYLFPLSPMYRLLYDDRDIFVYSORENRAELQRFVDEGT 120  
 QY 121 DGYEQFMEQERRKRFNALYPCITRDYSSLSKSFSLDLIKALWLAFFPKSVFNNLQGYFNOE 180  
 DB 121 DGYEQFMEQERRKRFNALYPCITRDYSSLSKSFSLDLIKALWLAFFPKSVFNNLQGYFNOE 180  
 QY 181 KRLAFPCQSKYLGHSPWECPALFTMLPYLHEHYGIYHVKGGLNRIAAAOVIAENGGE 240  
 DB 181 KRLAFPCQSKYLGHSPWECPALFTMLPYLHEHYGIYHVKGGLNRIAAAOVIAENGGE 240  
 QY 241 IHLNSEIESLIIENGAAKGVKLOHGAELRGDEVIINADFAAMTHLVKPGVLKXYTPENL 300  
 DB 241 IHLNSEIESLIIENGAAKGVKLOHGAELRGDEVIINADFAAMTHLVKPGVLKXYTPENL 300  
 QY 301 KOREYSCSTFMLYGLDKIYDLPHTTIVPAKDYTNIRNIENDKTLTDDFSPYVQNASAS 360  
 DB 301 KOREYSCSTFMLYGLDKIYDLPHTTIVPAKDYTNIRNIENDKTLTDDFSPYVQNASAS 360  
 QY 361 DDLAPAGKSAALYLVLPMPNNDGLDWAHCONVREQVLDTLGARLGSLDIRAHIECEKI 420  
 DB 361 DDLAPAGKSAALYLVLPMPNNDGLDWAHCONVREQVLDTLGARLGSLDIRAHIECEKI 420  
 QY 421 ITPTQWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEEELANCYLVGGTTHPGSGLPTIYE 480  
 DB 421 ITPTQWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEEELANCYLVGGTTHPGSGLPTIYE 480  
 QY 481 SARISAKLISQHRVRFKDIHAHSAWLKKAKA 511  
 DB 481 SARISAKLISQHRVRFKDIHAHSAWLKKAKA 511

## RESULT 4

ADAL4534  
 ID ADAL4534 standard; protein; 511 AA.

ADAL4534;

06-NOV-2003 (first entry)

Methylomonas sp. 16a CrtN protein SEQ ID NO:4.

carotenoid biosynthetic enzyme; CBE;

omega-aldehyde-functionalised carotenoid;

omega-carboxy-functionalised carotenoid; C30 carotenoid;

carotenoid biosynthesis; antioxidant; microbial; pigment; feed additive;

CrtN.

Methylomonas sp.

WO2003068917-A2.

21-AUG-2003.

11-FEB-2003; 2003WO-US004150.

11-FEB-2002; 2002US-0355939P.

(DUPO) DU PONT DE NEMOURS & CO E I.

Cheng Q, Norton KC, Tao L;

WPI; 2003-697524/66.

N-PSDB; ADAL4533.

New nucleic acid encoding carotenoid biosynthesis enzymes, useful for  
 preparing functionalized carotenoids, e.g. pigments and antioxidants,  
 also encoded polypeptides.

Example 4; Page 91-93; 125pp; English.

CC The present invention describes nucleic acid sequences encoding  
 CC carotenoid biosynthetic enzymes (CBEs). Also described: (1) a CBE  
 CC polypeptides; (2) a chimeric gene (CG) containing a CBE linked to  
 CC regulatory sequences; (3) host cell transformed with CG; (4) obtaining a  
 CC nucleic acid that encodes a CBE; (5) products of method (4); (6)  
 CC producing an omega-aldehyde-functionalised carotenoid (X); (7) producing  
 CC an omega-carboxy-functionalised carotenoid (Y); (8) producing C30  
 CC carotenoids (Z); and (9) regulating carotenoid biosynthesis in an  
 CC organism. CBE sequences have antioxidant activity. CBEs can be used for  
 CC microbial production of aldehyde- or carboxy-functionalised carotenoids  
 CC and C30 carotenoids, and are potentially useful as pigments and  
 CC antioxidants, e.g. as feed additives. The present sequence represents a  
 CC CrtN CBE protein isolated from *Methylobacterium* sp. strain 16a, which is used  
 CC in the exemplification of the present invention.  
 CC  
 CC Sequence 511 AA;  
 CC  
 CC Query Match 100.0%; Score 2711; DB 6; Length 511;  
 CC Best Local Similarity 100.0%; Pred. No. 6.4e-265;  
 CC Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC 2Y 1 MANTGHHIIVGAGCGGICAGMLLSORGFKVSIFDKHABIGGRNPINMGFTDGTGPTL 60  
 CC 2b 1 MANTGHHIIVGAGCGGICAGMLLSORGFKVSIFDKHABIGGRNPINMGFTDGTGPTL 60  
 CC  
 CC 2Y 61 LKGVLDLDFELCRRSEDEYLFPLSPMYRLLYDDRDIFVYSDRENRAELQRFDEGT 120  
 CC 2b 61 LKGVLDLDFELCRRSEDEYLFPLSPMYRLLYDDRDIFVYSDRENRAELQRFDEGT 120  
 CC  
 CC 2Y 121 DGTQFMQEKRRNALYPCITRDYSSLSKFLSLDLIKALPWLAPKSVNNIGQYFNOE 180  
 CC 2b 121 DGTQFMQEKRRNALYPCITRDYSSLSKFLSLDLIKALPWLAPKSVNNIGQYFNOE 180  
 CC  
 CC 2Y 181 KRLAFQSQYLVGNSPWCPCALFTMLPYLEHEYGIHVKGGLNRLAAMAQVIAENGGE 240  
 CC 2b 181 KRLAFQSQYLVGNSPWCPCALFTMLPYLEHEYGIHVKGGLNRLAAMAQVIAENGGE 240  
 CC  
 CC 2Y 241 IHLNSESIESLIENGAAGKVKLQHGABLRGDEVIIINADPAHAMTHLVKPVGLKXYTPENL 300  
 CC 2b 241 IHLNSESIESLIENGAAGKVKLQHGABLRGDEVIIINADPAHAMTHLVKPVGLKXYTPENL 300  
 CC  
 CC 2Y 301 KOREYSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNIFDNKTLIDDFSFFVQNASAS 360  
 CC 2b 301 KOREYSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNIFDNKTLIDDFSFFVQNASAS 360  
 CC  
 CC 2Y 361 DDLAPACKSALYLVPMPPNDSGLDQAHQCNVREQVLDTLGARGLSDIRAHIECEKI 420  
 CC 2b 361 DDLAPACKSALYLVPMPPNDSGLDQAHQCNVREQVLDTLGARGLSDIRAHIECEKI 420  
 CC  
 CC 2Y 421 ITPQWETDEHYKCATFSLSHKFSQMLYWRPHNRPFEELANCYLVGGTHPGSGLPITYE 480  
 CC 2b 421 ITPQWETDEHYKCATFSLSHKFSQMLYWRPHNRPFEELANCYLVGGTHPGSGLPITYE 480  
 CC  
 CC 2Y 481 SARISAKLISQHRVRFKDIASHAWLKAKA 511  
 CC 2b 481 SARISAKLISQHRVRFKDIASHAWLKAKA 511  
 CC  
 CC RESULT 5  
 CC ABB49224  
 CC ID ABB49224 standard; protein; 490 AA.  
 CC XX  
 CC AC ABB49224;  
 CC XX  
 CC XX 05-FEB-2002 (first entry)  
 CC DE  
 CC DE Listeria monocytogenes protein #1928.  
 CC XX  
 CC XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 CC XX vitamin B12; bacterial infection; disease.  
 CC OS  
 CC OS Listeria monocytogenes.

PN WO200177335-A2.  
 XX 18-OCT-2001.  
 XX 11-APR-2001; 2001WO-FR001118.  
 XX 11-APR-2000; 2000FR-00004629.  
 XX (INSP) INST PASTEUR.  
 XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Deboux P;  
 XX Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cosset P;  
 XX Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
 XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
 XX Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
 XX Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
 XX Madueno E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 XX Rose M, Voss H;  
 XX WPI; 2002-010914/01.  
 XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
 XX and prevention of *Listeria* and related bacterial infections, and related  
 XX polypeptides.  
 XX Claim 6; SEQ ID NO 1929; 192pp; French.  
 XX The present invention relates to the genome sequence of *Listeria*  
 XX monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
 XX it are useful for selecting probes and primers for detecting genes in *L.*  
 XX monocytogenes and related organisms, and for studying genetic  
 XX polymorphisms and other genomes. The present sequence is a protein  
 XX encoded by the genome sequence of the present invention. Proteins  
 XX expressed from the genome sequence are useful for raising specific  
 XX antibodies, identification of *L. monocytogenes* and related organisms, and  
 XX for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 XX B12. The genome sequence and proteins encoded by it are also useful for  
 XX selecting compounds that regulate gene expression and cell replication  
 XX and modulate *L. monocytogenes*-related diseases. In addition, the genome  
 XX sequence and proteins encoded by it are useful in pharmaceutical and  
 XX vaccine compositions for the treatment or prevention of infections by *L.*  
 XX monocytogenes and related organisms. Note: The sequence data for this  
 XX patent did not form part of the printed specification, but was obtained  
 XX in electronic format directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 490 AA;  
 XX  
 XX Query Match 34.3%; Score 929.5; DB 5; Length 490;  
 XX Best Local Similarity 37.6%; Pred. No. 1.5e-84;  
 XX Matches 184; Conservative 98; Mismatches 199; Indels 9; Gaps 1;  
 XX  
 XX 5 KHIIVGAGCGGICAGMLLSORGFKVSIFDKHABIGGRNPINMGFTDGTGPTL 64  
 XX 7 KKAIIGAGCGGICAGMLLSORGFKVSIFDKHABIGGRNPINMGFTDGTGPTL 66  
 XX  
 XX 65 VLDEMFCRRSEDEYLFPLSPMYRLLYDDRDIFVYSDRENRAELQRFDEGT 124  
 XX 67 ULTSFMDNENILYVSLPIHPIHTLYPKDITFPLYSQSETKAVITQYFPCEDGFD 126  
 XX  
 XX 125 QMBQERKRRNALYPCITRDYSSLSKFLSLDLIKALPWLAPKSVNNIGQYFNOE 184  
 XX 127 REMKENTKMLYISFLNQFNYSLLDFRPTTLRAIPSLMGRSLMDLSYFNSKYLRL 186  
 XX  
 XX 185 AFCPOSKYLGMSPWPCALFTMLPYLREHYGIHVKGGLNRLAAMAQVIAENGGE 244  
 XX 187 AFSLQWRVILGNSPWDIPAAYSIIFPSEYVYGFPIGGQNKIVAMQVVTENKGFEN 246  
 XX  
 XX 245 SPIESLIENGAAGKVKLQHGABLRGDEVIIINADPAHAMTHLVKPVGLKXYTPENL 304  
 XX 247 SEVTEFESNGKEITGAVLANGKTEADYITFLNDFIVSLTN-----EHPDLQTK 297  
 XX  
 XX 305 YSCSTFMYLGLDKIYDLPHHTIVFAKDYTNIRNIFDNKTLIDDFSFFVQNASAS 364

```

b 298 YSSAFILYGLTKVLPFSHQSIIPENYREFAHNTMHKILSKDIAIHLNPSATDNTM 357
y 365 APAGKSALVLPMPNDSGLDQAHQCNVRQVLDLTGALGLSDIRAHIECEKIITPQ 424
b 358 APINSSIRIMVPVNNNTSNIDKKETPAFRQLVLETWKERLEIPDLSQIEEYIITPI 417
y 425 TWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEELANCYLGGGTHPGSGGLTIYESARI 484
y 418 DWEKKYHVHKGAIPLQLHWRQHGVLHPSKSPKPNLFVIGAGAMSGSSLPFIENAIQI 477
y 485 SAKLISQKXER 494
y 478 ATQKFLQXEK 487

```

```

RESULT 6
B032531
D ABU32531 standard; protein; 490 AA.
D ABU32531;
K 19-JUN-2003 (first entry)
K Protein encoded by Prokaryotic essential gene #18058.
K Antisense; prokaryotic essential gene; cell proliferation; drug design.
K Listeria monocytogenes.
K WO200277183-A2.
K 03-OCT-2002.
K 21-MAR-2002; 2002WO-US009107.
K 21-MAR-2001; 2001US-00815242.
K 06-SEP-2001; 2001US-00948993.
K 25-OCT-2001; 2001US-0342923P.
K 08-FEB-2002; 2002US-00072851.
K 06-MAR-2002; 2002US-0362699P.
K (ELIT-) ELITRA PHARM INC.
K Wang L, Zamudio C, Malone C, Haselbeck R, Ohlseen XL, Zyskind JW,
K Wall D, T-awick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
K WPI; 2003-029926/02.
K N-PSDB; ACA36401.
K New antisense nucleic acids, useful for identifying proteins or screening
K for homologous nucleic acids required for cellular proliferation to
K isolate candidate molecules for rational drug discovery programs.
K Claim 25; SEQ ID NO 60455; 1766pp; English.

```

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a

```

CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 490 AA;

```

```

Query Match 34.3%; Score 929.5; DB 6; Length 490;
Best Local Similarity 37.6%; Pred. No. 1.5e-84;
Matches 184; Conservative 98; Mismatches 199; Indels 9; Gaps 1;

QY 5 KHIIVGAGPGGLCAGMLLSQSGFKVSIYDQKAEYGGNRINNGFTDGTGPTLLMKG 64
Db 7 KKIATIGAGPGGLAAGMLLSQSGYQVNTYERKDRIGRTAHRMGKYSFDVGPSTMTM 66
QY 65 VLDEMFELCERSESDYLEPLSPMYRLLYDDRDFVYSDRENMAELQRFDEGTGYE 124
Db 67 VLTSLFMDCNRLDYVSLPINPIHTLYFKDITFPLYSYDSETRAVIQTYPFGEDGFD 126
QY 125 QNMQQRKFNALYPCITRDYSLXSLFLSLDLIKALPWLAPPKSVFNNLQYFNQKRL 184
Db 127 RFMKENTKMLYISPLNQFNYSLSDFRPTTIRAPSLTGRSLMDLRSYFNISKYLL 186
QY 185 APCFQSKYLGNSPWECPALFTMLPYLEHYHYHVGGLNIAAAMAQVIAENGGEIHLN 244
Db 187 AFSLOWRYLGKSPWDIPRAYSIIIPSEYTYGTFHPIGGOKLIVEAMQVYVTKKPF 246
QY 245 STESLIIEGAAGKVGKLGQGAELRGDEVIINADPAHAMTHLVKPGVLKKYTPENLKORE 304
Db 247 SEVTEPESNGKEITGAVLANGKTIADYVFTNLDFTYSLTN-----EHPDKLDTKE 297
QY 305 YSCSTPMYVLGDKLYLDLPHHTIVFAKDYTTNIRNIFDNKLTDDPSFVQNASASD 364
Db 298 YSSAPFLYGLTKVLPFSHQSIIPENYREFAHNTMHKILSKDIAIHLNPSATDNTM 357
QY 365 APAGKSALVLPMPNDSGLDQAHQCNVRQVLDLTGALGLSDIRAHIECEKIITPQ 424
Db 358 APINSSIRIMVPVNNNTSNIDKKETPAFRQLVLETWKERLEIPDLSQIEEYIITPI 417
QY 425 TWETDEHYVKGATFSLSHKFSQMLYWRPHNRFEELANCYLGGGTHPGSGGLTIYESARI 484
Db 418 DWEKKYHVHKGAIPLQLHWRQHGVLHPSKSPKPNLFVIGAGAMSGSSLPFIENAIQI 477
QY 485 SAKLISQKXER 494
Db 478 ATQKFLQXEK 487

```

```

RESULT 7
AAR64269
ID AAR64269 standard; protein; 499 AA.
XX AAR64269;
AC AAR64269;
XX 16-OCT-2003 (revised)
DT 27-AUG-2003 (revised)
DT 02-OCT-1995 (first entry)
XX Anabaena epsilon-carotene desaturase protein.
XX Anabaena, epsilon-carotene desaturase gene; SAN380H; J852; plant;
KW phytoene desaturase inhibiting herbicide; norflurazon; fluridone;
KW Brinia uredovora; lycopene converting activity; transit peptide.

```

XX DS Nostoc sp: "PCC 7120".  
XX Key Location/Qualifiers  
XX Misc-difference 296 /note= "encoded by ATC"  
XX Misc-difference 299 /note= "encoded by TCA"  
XX Misc-difference 300 /note= "encoded by GAC"  
XX Misc-difference 302 /note= "encoded by CTT"  
XX Misc-difference 303 /note= "encoded by TAT"  
XX Misc-difference 304 /note= "encoded by TTG"  
XX Misc-difference 305 /note= "encoded by GGT"  
XX Misc-difference 306 /note= "encoded by ATC"  
XX Misc-difference 307 /note= "encoded by AAT"  
XX Misc-difference 308 /note= "encoded by CGC"  
XX Misc-difference 309 /note= "encoded by CGC"  
XX Misc-difference 310 /note= "encoded by TAC"  
XX Misc-difference 311 /note= "encoded by GAA"  
XX Misc-difference 312 /note= "encoded by GAT"  
XX Misc-difference 313 /note= "encoded by TTA"  
XX Misc-difference 314 /note= "encoded by CCT"  
XX Misc-difference 315 /note= "encoded by CAT"  
XX Misc-difference 316 /note= "encoded by CAT"  
XX Misc-difference 317 /note= "encoded by CAA"  
XX JP06343473-A.  
XX 20-DEC-1994.  
XX 08-JUN-1993; 93JP-00163926.  
XX 08-JUN-1993; 93JP-00163926.  
XX (KIRI ) KIRIN BREWERY KK.  
XX WPI: 1995-069305/10.  
XX N-PSDB; AAQ80736.  
XX Prodn of bleaching herbicide-resistant plants - by transformation with  
XX the Erwinia phytoene desaturase gene.  
XX Example 9; Page 11-13; 20pp; Japanese.  
XX The sequence of the Anabaena strain PCC7120 epsilon-carotene desaturase  
XX protein. The corresponding gene was used in a method for the production  
XX of plants resistant to epsilon-carotene desaturase inhibiting herbicides  
XX e.g. SAN380H, J852, etc., or phytoene desaturase inhibiting herbicides  
XX e.g. norflurazon, fluridone, etc., by introducing into the plant a DNA  
XX sequence that encodes a polypeptide, preferably the Erwinia-derived  
XX phytoene desaturase e.g. the E.uredovora crtI gene, or a variant e.g. the  
XX Anabaena sequence given in the specification, having a phytoene to  
XX lycopene converting activity. The DNA sequence also contains a sequence  
XX coding for a transit peptide. (Updated on 27-AUG-2003 to correct OS  
XX field.) (Updated on 16-OCT-2003 to standardise OS field)

SQ Sequence 499 AA;  
Query Match 28.9%; Score 783.5; DB 2; Length 499;  
Best Local Similarity 35.0%; Pred. No. 9.2e-70;  
Matches 175; Conservative 103; Mismatches 201; Indels 21; Caps 10;  
QY 4 TKHIIYVAGGGGICAGMLLSQRFKYSIFDKHAEITGGRRPINMNGFTDPTGPTLLMK 63  
DB 2 SKKVAIVAGGGGATATRLAGLQIQVEIFEAERVGGRRGGEVSYADFTGPTLIQLP 61  
QY 64 GVLDEMPCLCERSELYLEFLPLSPMYRL-LYDRIIFVYSRENMRAEIQRVDETDG 122  
DB 62 HLYKELFEELAGNFADYVQLKRLPEYTRLAKFWDTQDIDTDLQSPKTLATLRSDLPLA 121  
QY 123 YEQFMEQERKFNALY-PCITRDYSSLSKSLSL-DLIKAL---PWLAFPKSVNNLQYF 177  
DB 122 FDRWYSBHIRKYELGYKPYLAGPARSIFGYLRPDDLKMLSLSPFW----ENLQHFWRFP 177  
QY 178 NOEKRLAFCQSKYLQMSWPCPALFTMLPYLEHEXYGYHVKGLNRIAAAOVIAEN 237  
DB 178 QDERLVYDLRYPSKYLGMHPTVASSVSLIPFLFESQGVNHPVGGFRALAAQGLANAQDL 237  
QY 238 GGBHLNSETESLIENGAAKGVKLOHGAELRGDEVTLNADFAHAMTHLVKPGVLKKTYP 297  
DB 238 GVKIHLSPVHQIWIWQGVRGLELADASHQFDTVVINADFAFAYVHLLPTTSARGYTD 297  
QY 298 ENLKOREYSCSTFMLYLGLDKIY-DLPHRTIIVFAKDYTTNIRNI----FDNKLTLD-DPS 351  
DB 298 NKLGMQVFCSTFMLYLGINERYEDLPHQIYL----SDMIRLERPWVDDSDALDETDP 353  
QY 352 FYVONASASDOSLAPAGKSAIYLVPMNDSDGLDQWQACQNYREQVLDTLGARLGSLDI 411  
DB 354 FYVCNPTIIDPSNAPAGHSTLFVLVPIPTNSYAVDWDIKQKSYTDFILKRLHL-LGVHNI 412  
QY 412 RAHTECEKIITPQTWETDEHVYKGAFTSLSHKFSQMLYEPHNRFEELANCYLWGGGTHP 471  
DB 413 EKHIVTQCYTAQSWLDDYRVHLGAVFNLSHNLTLQLGPPRPPIRSENIAGLYWIGGAVHP 472  
QY 472 GSGLPITYESARISAKLISQ 491  
DB 473 GSGLLTILEASRSAAAGFTHQ 492  
RESULT 8  
ABW70123  
ID ABW70123 standard; protein; 494 AA.  
XX AC ABW70123;  
XX DT 20-NOV-2003 (first entry)  
XX DE Photorhabdus luminescens protein sequence #3220.  
XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
XX KM detection; food; gene expression; plant; animal; microorganism; toxin;  
XX KM antibiotic; biopesticide; virulence factor; disease model; plague;  
XX KM whooping cough.  
XX OS Photorhabdus luminescens.  
XX PN WC200294867-A2.  
XX PD 28-NOV-2002.  
XX PF 07-FEB-2002; 2002NO-IB003040.  
XX PR 07-FEB-2001; 2001FR-00001659.  
XX (INSP ) INST PASTEUR.  
XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX XX Duchaud E, Tsourit S, Glaser P, Frangeul L, Kunst P, Danchin A;  
PI Buchrieser C;

XX WPI; 2003-148459/14.  
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
XX  
XX Claim 2; SEQ ID NO 3220; 1205pp; French.  
XX  
XX The invention relates to the isolation of genes and their encoded  
XX proteins from Photorhabdus luminescens. The isolated sequences are  
XX sources of probes and primers for detecting the genome of P. luminescens  
XX and related species; to study polymorphisms; for gene analysis and for  
XX detection/amplification of the genes. Antibodies (Ab) raised against the  
XX polypeptides encoded by the genes are used for detection/identification  
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
XX carry a gene-containing vector are used to select compounds that  
XX modulate, regulate, induce or inhibit expression of the genes in plants,  
XX animals or microorganisms other than P. luminescens and are able to alter  
XX response or sensitivity to toxins and antibiotics produced by P.  
XX luminescens. Cells transformed to express the genes are useful for  
XX recombinant production of the proteins, particularly toxins and  
XX antibacterials useful as insecticides, bactericides and fungicides. The  
XX genes, proteins, vectors containing the genes and Ab are also useful  
XX therapeutically (to treat microbial infection by bacteria or fungi that  
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
XX biopesticides. Other uses of the genes and the proteins are as virulence  
XX factors and for identifying targets of human diseases for which P.  
XX luminescens is a model (particularly plague and whooping cough). This  
XX sequence represents one of the isolated P. luminescens proteins  
XX  
XX Sequence 494 AA;

Query Match 26.3%; Score 713; DB 6; Length 494;  
Best Local Similarity 31.4%; Pred. No. 1.2e-62;  
Matches 159; Conservative 109; Mismatches 212; Indels 24; Gaps 10;  
Y 8 IIVGAPGGICAGMLLSQSGFKVSIIDKHAETGGRNRPINMNGFTDGTGPTFLMKGVLD 67  
b 5 LVTCAGFGGLALAIRQAGIPTCELEQDKPGRAYVYKEQGTFTDAGPTVITAPNVE 64  
Y 68 EMPELCERSEDEYLEPLPSPMYRLYDDRDIFVY-SDRENRAELQRFVDBGTGYSQF 126  
b 65 ELTQAGKRWADYDVLPHVPFVRLCWESGKFDYDNDQOHLAQIHTFNPRDVGNYRFP 124  
Y 127 MEGERKFNALYPCI-TRDYSSLKSLSL--DLIKALPWLAPPKSVFNNLGOYFNQEKMR 183  
b 125 LDYSREAFNEGKLGKTVPLSPFSLDMLSAAPOLIRLHAW---RSVYQVAFIKDESIR 180  
Y 184 LAFCSQKYLGSWPCPALFTMLPYLEHEGYIYHVGGLARIAMAQVIAENGGEIHL 243  
b 181 QAFSFLSLIGGNPFAASIIYLIHALREWGVPFGTSAVLEAMVKLFTDGGIEEL 240  
Y 244 NSBIESLIIENGAAGKVLQHGAEIAGDEVIINADPAHAMTHLV---KPGVLKXYTPENL 300  
b 241 NAKVKFTTHGNRTVCGVLADGENMTCDVVASNADVIHTYKHLGQHPVGIART---L 297  
Y 301 KQREYSCSTFMGLYGLDKYI-DLPHHTIVFAKDYTNIRNIPDNKTLTDDTSFYVQNAS 359  
b 298 VRKMSNSFLVYFGLNHHHTQLAHTVCGPGRYKELIEDIFIDRLHSEDPSLYLHAPSV 357  
Y 360 SDDSLAPAGKALYVLVMPEN-NDSLDMQACQNVREOVLDTLGLRGLSDIRAHBCE 418  
b 358 TDSLAPRCASYYVLAPVPHLGTANLWIDEGPRIDRIIPAYL-EKTYMGLGLKQLVYH 416  
Y 419 KIITPTWTDEHYKAGATFSLSHKFSQMLYWRPNRFEELANCYLVGGGTHPGSLPTI 478  
b 417 RIPTPTDFDQLNAHLGSAFSPPELLTQSAWPFPHNRNDRNIDNLYLVAGTHPGAGIPGV 476  
Y 479 YESARISAKLISKQHVRFKDTA 501  
b 477 IGSKATATLM-----LEDIA 492

RESULT 9  
AAE22314  
ID AAE22314 standard; protein; 492 AA.  
XX  
XX AC AAE22314;  
XX  
XX DT 25-JUL-2002 (first entry)  
XX  
XX DE Pantoea stewartii phytoene desaturase (CrtL) enzyme.  
XX  
XX KW Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;  
XX anti-oxidant; steroid; flavour; fragrance; electro-optic application;  
XX aquaculture; enzyme; phytoene desaturase; CrtL.  
XX  
XX OS Pantoea stewartii.  
XX  
XX FM WO200218617-A2.  
XX  
XX PD 07-MAR-2002.  
XX  
XX PF 04-SEP-2001; 2001WO-US027420.  
XX  
XX PR 01-SEP-2000; 2000US-0229858P.  
XX  
XX PA 01-SEP-2000; 2000US-0229907P.  
XX (DUPO) DU PONT DE NEMOURS & CO E I.  
XX  
XX PI Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;  
XX PI Odom JM, Picataggio SK, Rouviere PE;  
XX  
XX DR WPI; 2002-351711/38.  
XX  
XX DR N-PSDB; AAD35512.  
XX  
XX PT Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by  
XX using microorganisms having a nucleic acid molecule encoding enzymes in  
XX the carotenoid biosynthetic pathway and which metabolize single carbon  
XX substrates.  
XX  
XX PS Claim 19; Page 141-143; 156pp; English.  
XX  
XX CC The invention relates to a method for producing carotenoid compounds. The  
XX method comprises a transformed metabolising host cell, comprising  
XX suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule  
XX encoding an enzyme in the carotenoid biosynthetic pathway, under the  
XX control of regulatory sequences, and contacting the host cell with carbon  
XX substrate to produce a carotenoid compound. The method is useful for  
XX producing carotenoid compounds such as antheraxanthin and astaxanthin, by  
XX using microorganism having a nucleic acid molecule encoding enzymes in  
XX the carotenoid biosynthetic pathway and which metabolise single carbon  
XX substrates. The carotenoids have potent anti-oxidant properties useful in  
XX diet, and aquaculture elements. The carotenoids are also useful as  
XX intermediates in the synthesis of steroids, flavours and fragrances and  
XX compounds for potential electro-optic applications. The present sequence  
XX is Pantoea stewartii phytoene desaturase (CrtL) enzyme used in the  
XX invention  
XX  
XX SQ Sequence 492 AA;  
Query Match 25.9%; Score 703.5; DB 5; Length 492;  
Best Local Similarity 31.7%; Pred. No. 1.1e-61;  
Matches 159; Conservative 105; Mismatches 210; Indels 25; Gaps 10;  
QY 5 KHIIIVGAPGGICAGMLLSQSGFKVSIIDKHAETGGRNRPINMNGFTDGTGPTFLMKG 64  
Db 2 KPTVTVGAGFGGLALAIRQAGIPTCELEQDKPGRAYVYKEQGTFTDAGPTVITDPS 61  
QY 65 VLDEMPCLERSEDEYLEPLPSPMYRLYDDRDIFVY-SDRENRAELQRFVDBGTGYSQF 123  
Db 62 AIEELPALAGKQLKQYVELLPVTPFVRLCWESGKFDYDNDQOHLAQIHTFNPRDVG 121  
QY 124 EQPMEQERKFNALYPCI-TRDYSSLK-----DLIKALPWLAPPKSVFNNLGOYFN 175  
Db 122 RAFLDYSAVFNEGY-----LKGTVPLSPFSLDMLSAAPOLIRLHAW---RSVYQVAFIK 172



QY 176 YNQEKMELAFQSKYLGMSFWPCALFTMLPYLEHSEYGIYHVKGGLNRIAAQAQVIA 235  
DB 173 YIEDEHLQAQSFHSLVGGNPFATSSYITLIHALEREWGVPFRGGTGALVNGMIKLFQ 232  
QY 236 ENGGEIHLNIESIESLIIENGGAAGVYKLOHGAELRGDEVIINADFAHAMTHLVK--PGVLK 293  
DB 233 DLGGEVILNARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSQHPAAK 292  
QY 294 KYTPENLKQREYSCSTFMYLGLDKIYD-LPHHTIVFAKYTTNIRNIFDNKTLTDDFSF 352  
DB 293 Q--AKKQSKMSNSLFVLYFGLNHHHDQLAHTVCFGRYRELIHEIFNHDGLAEDFSL 350  
QY 353 YVONASASDDSLAPAGKSAIYLVPMEN-NDSGLDWQAHCONVREQVLDITLGARLGLSDI 411  
DB 351 YLHAPCVTDPSLAPEGCGSYVYLAPEVHLGTANLDWAVEGPRLDRIYDYLEQHY-MPGL 409  
QY 412 RAHIECEKIITPQWETDEHYKATFSLSHKFSQMLYWRPHNRPBELANCYLVGCGTHP 471  
DB 410 RSQLVTHMFTHFFDRDELNAWQSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHP 469  
QY 472 GSGLPFTIVESARISAKLI 489  
DB 470 GAGIPGVIGSAKATAGLM 487

## RESULT 10

QY AAO16021 ID AAO16021 standard; protein; 492 AA.

AC AAO16021;

DT 20-FEB-2003 (first entry)

DE Pantoea stewartii phytoene desaturase.

XX Carotenoid; crt.

XX Pantoea stewartii.

XX WO200279395-A2.

XX 10-OCT-2002.

XX 25-JAN-2002; 2002WO-US002124.

XX 26-JAN-2001; 2001US-0264329P.

XX 04-MAY-2001; 2001US-0288984P.

XX (CRGI ) CARGILL INC.

XX De Souza ML, Kollmann SR, May CA, Schroeder WA;

XX WPI: 2003-075455/07.

XX N-PSDB; ABT14193.

XX Novel isolated nucleic acid useful e.g. to engineer host cells with the

XX ability to produce particular carotenoids and polypeptides useful in cell

XX -free systems to make particular carotenoids.

XX Claim 26; Page 63-64; 74pp; English.

XX The invention comprises the amino acid and coding sequence of a number of

XX carotenoid (crt)-related proteins. The crt-related DNA and protein

XX sequences of the invention are useful for engineering cells which are

XX able to produce carotenoids. The present amino acid sequence represents a

XX crt-related protein of the invention

XX Sequence 492 AA;

XX Query Match 25.9%; Score 703.5; DB 6; Length 492;

XX Best Local Similarity 31.7%; Pred. No. 1.1e-61;

XX Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;

QY 5 KHIITVAGPGGLCAGMLLSQSGPKVSIQFKHAIQGRNPPINNGGTFDTGPTFLLMKG 64  
DB 2 KPTTIVAGPGGLALAIRLQAAGIPVLLLEQDRPGGRVAVYQCGQTFDAGPTVITDPS 61  
QY 65 VLDEFELCERSESDYLBFLPLSPMYRLLYDDRIDFVY-SDRENMAEALORVFEDETDGY 123  
DB 62 ALEELFALAGKQLKDYVELLPVTFYRLCWESGKVFNVDADQALEAQIQQNFEDVAGY 121  
QY 124 BQMEQEKREFNALYPCITRDYSSLK-----SFLSL-DLHKALPWLA---FPKSVFNNLQ 175  
DB 122 RAPLDYSRAVNEG-----LKLGTVPFLSFKXDMLRAPQALAKLQAWRSVYSKVAG 172  
QY 176 YNQEKMELAFQSKYLGMSFWPCALFTMLPYLEHSEYGIYHVKGGLNRIAAQAQVIA 235  
DB 173 YIEDEHLQAQSFHSLVGGNPFATSSYITLIHALEREWGVPFRGGTGALVNGMIKLFQ 232  
QY 236 ENGGEIHLNIESIESLIIENGGAAGVYKLOHGAELRGDEVIINADFAHAMTHLVK--PGVLK 293  
DB 233 DLGGEVILNARVSHMETVGDKIQAQVLEDGRFRFETCAVASNADVVHTYRDLSSQHPAAK 292  
QY 294 KYTPENLKQREYSCSTFMYLGLDKIYD-LPHHTIVFAKYTTNIRNIFDNKTLTDDFSF 352  
DB 293 Q--AKKQSKMSNSLFVLYFGLNHHHDQLAHTVCFGRYRELIHEIFNHDGLAEDFSL 350  
QY 353 YVONASASDDSLAPAGKSAIYLVPMEN-NDSGLDWQAHCONVREQVLDITLGARLGLSDI 411  
DB 351 YLHAPCVTDPSLAPEGCGSYVYLAPEVHLGTANLDWAVEGPRLDRIYDYLEQHY-MPGL 409  
QY 412 RAHIECEKIITPQWETDEHYKATFSLSHKFSQMLYWRPHNRPBELANCYLVGCGTHP 471  
DB 410 RSQLVTHMFTHFFDRDELNAWQSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHP 469  
QY 472 GSGLPFTIVESARISAKLI 489  
DB 470 GAGIPGVIGSAKATAGLM 487

## RESULT 11

ABP96688

ID ABP96688 standard; protein; 492 AA.

XX AC ABP96688;

XX 03-JUN-2003 (first entry)

XX Pantoea stewartii phytoene desaturase SEQ ID NO:6.

XX Pantoea stewartii; carotenoid biosynthetic enzyme; crtE; crtX; crtY;

XX crtI; crtB; crtZ; phytoene desaturase; enzyme; phytoene; carotenoid.

XX Pantoea stewartii.

XX WO2003016503-A2.

XX 27-FEB-2003.

XX 15-AUG-2002; 2002WO-US026647.

XX 15-AUG-2001; 2001US-0312646P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Brzostowicz PC, Cheng Q, Picataggio SK, Rouviere PS;

XX WPI; 2003-268323/26.

XX N-PSDB; ACC44762.

XX Novel nucleic acid molecule isolated from Pantoea stewartii encoding a

XX carotenoid biosynthetic enzyme, useful for regulating carotenoid

XX biosynthesis in an organism.

XX Claim 4; Page 63-65; 68pp; English.

X The present invention describes Pantoea stewartii carotenoid biosynthetic  
 C enzymes (I). More specifically described are the geranylgeranyl  
 C pyrophosphate synthase (crtE), zeaxanthin glucosyl transferase (crtX),  
 C lycopene cyclase (crtY), phytoene desaturase (crtI), phytoene synthase  
 C (crtB) and beta-carotene hydroxylase (crtZ) enzymes (see ABP9685 to  
 C ABP9690) encoded by ACQ44759 to ACQ44764. (I) can be used for regulating  
 C carotenoid biosynthesis in an organism, by over-expressing (I) in an  
 C organism, such that the carotenoid biosynthesis is altered in the  
 C organism. (I) and the genes encoding (I) are useful for converting  
 C phytoene to the carotenoids, for creating recombinant organisms that have  
 C the ability to produce various carotenoid compounds, and also for  
 C enhancing or manipulating carotenoid compounds. (I) can also be used for  
 C producing gene products having enhanced or altered activity  
 X  
 Q Sequence 492 AA;  
 Query Match 25.9%; Score 703.5; DB 6; Length 492;  
 Best Local Similarity 31.7%; Pred. No. 1.1e-61;  
 Matches 158; Conservative 105; Mismatches 210; Indels 25; Gaps 10;  
 Y 5 XHIIIVGAGFGGLCAGMLLSQSGFKVSIFDKHAEIGGRNRPINMNGFTDGTGPTFLMKG 64  
 b 2 KPTTVIGAGFGGLAIRLOAGIPVLLLEQDKGGRAYVYQEGFTDAGTVIDPS 61  
 Y 65 VLDEMFELCERSEDYLFPLSPMYRLLYDDRDIPVY-SDENNRALQRFVDEBGTGY 123  
 b 62 AIEELPALAGKQLKDYVELLPVTPFYRLCWESGKVFYNDQAEQIQQFNPRDVAGY 121  
 Y 124 EOPMEQERKRNALYPCITRDYSSLK---SFLSL-DLIKALPMLA---FPKSVFNLLGQ 175  
 b 122 RAFLDYSRAVFNQY-----LKLGVFPLSPFKDMLRAAPQLAKLQARSVYSKVAG 172  
 Y 176 YFNOEKRLAFQFQSKYLGMSWPECALFTMLPYLEHEYGIYHVKGGLNRJAAMAQVIA 235  
 b 173 YIDEHLRQAFSPHLLVGNPNFATSSIVTLTHALERWGVWFPFRGGTGAIVNGMIKLFQ 232  
 Y 236 ENGCEHIANSEIESLIENGAAGKWLQHGAEIRGDEVIIINADFAHATHLVK--PGVLK 293  
 b 233 DLUGGEVLNARVSHMETVGDKIQAQVLEDDRRFETCAVASNADVVHTYRDLLSQHPAAK 292  
 Y 294 KVTPEMLKQREYSCSTFMYLGLDKIYD-LPHHTIVFAKDYTNIRNIFDNKLTDDFSF 352  
 b 293 Q--AKKLQSKRNSLSFLVLYFGLNHHQDLAHTVCFGRYELIHEIFNHDGLAEDFSL 350  
 Y 353 YVONASASDSDSLAPAGKSALYLVVPMNP-NDSGLDQWQHCQNRQVLDLTGARLGLSDI 411  
 b 351 YLHAPCVTDPSLAPEGCGSYVVLAPVPHLGTANLDWAVEGPELRDRIIDYLEQHY-MEGL 409  
 Y 412 RAHIECEKIITPQTWETDEHYKGFATSLSHKFSQMLYWRPHNRPEELANCVLGGGTHP 471  
 b 410 RSQLVTHRMFTDPDFDELNAQGSAPSVPEILTQSAMFRFENRDKHIDNLYLVGAGTHP 469  
 Y 472 GSGLPITYSARISAKLI 489  
 b 470 GAGIPGVIGSAKATAGLM 487  
 RESULT 12  
 D BM72613  
 X ABM72613 standard; protein; 502 AA.  
 X  
 X ABM72613;  
 X  
 X 20-NOV-2003 (first entry)  
 X  
 X Staphylococcus aureus protein #1853.  
 X  
 W Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;  
 W enzymatic assay; antibiotic target.  
 X  
 X Staphylococcus aureus.  
 X

PN WC200294868-A2.  
 PD 28-NOV-2002.  
 XX 27-MAR-2002; 2002WO-IB002637.  
 PF 27-MAR-2001; 2001GB-00007661.  
 PR (CHIR-) CHIRON SPA.  
 XX Masignani V, Mora M, Scarselli M;  
 XX WPI; 2003-120786/11.  
 DR N-PSDB; AC741173.  
 XX New Staphylococcus aureus protein, useful as a vaccine for treating or  
 PF preventing Staphylococcal infection, specifically an infection caused by  
 PT S. aureus, e.g. sepsis.  
 XX Claim 1; SEQ ID NO 3706; 49pp; English.  
 PS  
 XX The invention relates to novel genes and encoded proteins from  
 CC Staphylococcus aureus. A composition comprising the S. aureus protein, a  
 CC nucleic acid encoding the protein, or an antibody to the protein, is  
 CC useful as a pharmaceutical, particularly as a vaccine for treating or  
 CC preventing infection due to Staphylococcus bacteria, specifically an  
 CC infection caused by S. aureus. The composition is particularly useful for  
 CC treating or preventing sepsis in a patient. The composition can also be  
 CC used for diagnostics. The protein is also used in an assay for enzymatic  
 CC studies and as a target for antibiotics. This sequence represents one of  
 CC the novel S. aureus proteins of the invention  
 XX  
 Q Sequence 502 AA;  
 Query Match 25.9%; Score 702.5; DB 6; Length 502;  
 Best Local Similarity 31.5%; Pred. No. 1.5e-61;  
 Matches 159; Conservative 108; Mismatches 196; Indels 41; Gaps 12;  
 QY 7 IIVGAGPGLCAGMLLSQSGFKVSIFDKHAEIGGRNRPINMNGFTDGTGPTFLMKGV 66  
 DB 3 IAVIGAGVTLGAAAARIASQGHVETFEKNNVGGRNQKKGDFTFDGMPTIVMPDVY 62  
 QY 67 DEMFELCERSEDYLFPLSPMYRLLYDDRD-IFVYSDENNRALQRFVDEBGTGYEQ 125  
 DB 63 KDVPYTAGKNYEDYIELRLQYIYDVYDHDRIITVPTDLAELQOMLESIEPGSTHGFMS 122  
 QY 126 FMEQERKRNALYPCITR-----DYSSLXSPSLDLIKALPMLAPKSVFNLL 173  
 DB 123 FLTDVYKYE-----IARYFLERTYRKPSDFYNTSLVCGAKIKTL-----NHA 167  
 QY 174 GQ----YFNOEKRLAFQFQSKYLGMSWPECALFTMLPYLEHEYGIYHVKGGLNRJAAA 229  
 DB 168 DQLEHYIDNEKIQLLAFQTLVIGIDPKRGSLSYSIIEMIEWMGVHFIKGMYGMAQG 227  
 QY 230 MAQVIAENGGEIHLNSETIESLIENGAAGKWLQHGAEIRG-DEVIINADFAHATHLVK 288  
 DB 228 LAQLNKDLGVNIELNABIEQIIIDPKFEADAIVKNGDIRKEDKILCTADPFPSVAESLMP 287  
 QY 289 P-CVLKAYTPENLKQREYSCSTFMYLGLD-KIYD-LPHHTIVFAKDYTNIRNIFDNKT 345  
 DB 288 DPAPIKKYPPHKIADLDYSCSAFLMTYIGIDIVTQVRLNLFVPSDDFGNIEEIFEGR- 346  
 QY 346 LTDDFSFYVONASASDSDSLAPAGKSALYLVVPMNP--NNDSGLDW--QAHCQNRQVLD 401  
 DB 347 LSYDPSIYVVVPAVADKSLAPAGKTYIYVLMPTPELKTGSGIDWSEALTOQIKELIYRK 406  
 QY 402 LGARLGLSDIRAHIECEKIITPQTWETDEHYKGFATSLSHKFSQMLYWRPHNRPEELAN 461  
 DB 407 LATIEVFEDIKSHIVSETIFTDNDPEQTHAKFGSAGLMPFLAQSNTYRPNQVSDYKD 466  
 QY 462 CYLVGGGTHPGSGLPITYESARIS 485  
 DB 467 LYFAGASTHPGAGVPILTSAKIT 490

QY	174	GO-----YFNQEKARLAF	CFQSKYLHSGSPNECPAL	FTMLPVLREHYGCIYHVKGGLNR	IAAA	227
DB	168	DQLEHYDINDEKIQKLAF	OTLYIGIDPKRGPSLSYII	IPRIEMFGVHFPIKGMVGMAGG		227
QY	230	MAQVIAENGGEIHLN	SEISLIIENCAAGVKVLOHGA	ELFG-DEVIINADFAHAMTHLVK		288
DB	228	LAQLNKDLGVNIELNA	SEIQIIDIIPKFRADA	IKVNGDIRKFDKILCTADFP	SVASLSMP	287
QY	289	P-GVLKKYTPENL	KOREYSGSTFMLVIGLD-KIYD-LPH	LTIVFAKDYTTNIRNIFDNKT		345
DB	288	DFADIKYTPHKAD	LDYCSAFMLTIGIDIDVTQVRLEN	VFESDDFRGNIEEIEGR-		346
QY	346	LTDDFSFYQNASAD	SDSLAPAGKSALYVLVMP--NND	SGLDW--QARCONVR	EQVLD	401
DB	347	LSYDPSIYVYVPA	VAADKSLAPEGKTLGIYVLMP	TPPELKTGEGIDWSDEALTQ	QIKSIIYRK	406
QY	402	LGARLGISDI	PAIECEKILTPQTNWTDENVYKAT	PSLCHKESQMLYWRPNR	FEELAN	461
DB	407	LATIEVPEDIKSH	IVSETITFPNDFQTYHAKGSA	FGMLPTLAOSNYRPNQ	SRDYKD	466
QY	462	CYLVGGGTHPGSG	LPTIYESARIS	485		
DB	467	LYFAGASTHPCAG	VPVILTSAKIT	490		
RESULT 14						
QY	AAR07466	ID AAR07466 standard; protein; 492 AA.				
XX	AAR07466;					
DT	24-OCT-2003	(revised)				
DT	28-JAN-1991	(first entry)				
XX	DE	Polypeptide with enzymatic activity for the conversion of phytoene into lycopene.				
XX	DE	Carotenoid biosynthesis; vitamin A; cancer; food coloring.				
XX	KW	Pantoea ananatis.				
OS	XX	EP393690-A.				
PN	XX	24-OCT-1990.				
PD	XX	20-APR-1990; 90EP-00107493.				
PF	XX	21-APR-1989; 89JP-00103078.				
XX	PR	05-MAR-1990; 90JP-00053255.				
XX	XX	(KIRI ) KIRIN BEER KK.				
XX	PI	Misawa N, Kobayashi K, Nakamura K;				
PI	XX	WPI; 1990-322212/43.				
DR	XX	N-PSDB; AAQ06296.				
PT	XX	DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn. of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.				
PT	XX	Claim 4; Fig 4; 40pp; English.				
PS	XX	Gene products are useful for the synthesis of carotenoids, useful as food coloring, vitamin A precursor, and possibly in prevention of cancer. See also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)				
CC	XX	Sequence 492 AA;				
CC	XX	SQ				

DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn. of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.

Claim 4; Fig 4; 40pp; English.

Gene products are useful for the synthesis of carotenoids, useful as food coloring, vitamin A precursor, and possibly in prevention of cancer. See also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)

Sequence 492 AA;

Query Match 25.8%; Score 700.5; DB 2; Length 492;  
Best Local Similarity 31.5%; Pred. No. 2.3e-61;  
Matches 157; Conservative 103; Mismatches 213; Indels 25; Gaps



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

! protein - protein search, using sw model

in on: February 29, 2004, 14:33:49 ; Search time 24.4213 Seconds  
(without alignments)  
3837.172 Million cell updates/sec

title: US-09-941-947A-20

sequence: 1 MSKUKAYLTVCQVERALD.....FGSEADLLRLSLYIERTH 237

scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 1017041 seqs, 315518202 residues

total number of hits satisfying chosen parameters: 1017041

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

SPTREMBL\_25.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_plant.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriaph.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	831	55.0	295	16 Q9HWY4	Q9HWY4 pseudomonas
2	811	53.7	295	16 Q8BQG6	Q8BQG6 pseudomonas
3	800.5	53.0	300	2 Q9S5P0	Q9S5P0 synchococc
4	791	52.4	295	16 Q8B9Q0	Q8B9Q0 pseudomonas
5	787.5	52.2	297	16 Q8ZVD4	Q8ZVD4 nitrosomona
6	729	48.3	298	16 Q9USM0	Q9USM0 neisseria m
7	715	47.4	231	16 Q8P7L1	Q8P7L1 xanthomonas
8	707	46.8	298	16 Q8DFA4	Q8DFA4 vibrio vuln
9	706	46.8	291	16 Q8PIY3	Q8PIY3 xanthomonas
10	705.5	46.7	298	16 Q8ZRD0	Q8ZRD0 vibrio para
11	705.5	46.7	298	16 Q8ZRD0	Q8ZRD0 salmonella
12	704	46.6	294	16 Q8ZXL2	Q8ZXL2 vibrio chol
13	704	46.6	294	16 Q8ZXL2	Q8ZXL2 vibrio chol
14	702.5	46.5	306	16 Q8ZC46	Q8ZC46 yersinia pe
15	695.5	46.1	299	16 Q8XE75	Q8XE75 escherichia
16	695.5	46.1	308	16 Q7VLU1	Q7VLU1 haemophilus

#### ALIGNMENTS

#### RESULT 1

Q9HWY4 Q9HWY4 PRELIMINARY; PRT; 295 AA.  
ID Q9HWY4  
AC Q9HWY4;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Geranyltransferase.  
GN IS OR PA0403.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
CX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
Garber R.L., Goltry L., Tolettino E., Westbrook-Wadman S., Yuan Y.,  
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen."  
RT Nature 406:959-964 (2000).  
DR EMBL; AB004821; AAC07430.1; --  
DR PIR; F83139;  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.  
DR InterPro; IPR000092; Polyprenyl\_synth.  
DR InterPro; IPR008949; Terpenoid\_synth.  
DR Pfam; PF00348; polyprenyl synth; 1.  
DR PROSITE; PS00723; POLYPRENYL SYNTHET\_1; 1.  
DR PROSITE; PS00444; POLYPRENYL SYNTHET\_2; 1.  
KW Transferase; Complete proteome.  
SQ SEQUENCE 295 AA; 31464 MW; EFDFFA6D66F5C126 CRC64;

Query Match

55.0%; Score 831; DB 16; Length 295;



```
01-JUN-2003 (TREMELrel. 24, Created)
01-JUN-2003 (TREMELrel. 24, Last sequence update)
01-OCT-2003 (TREMELrel. 25, Last annotation update)
Geranyltransferase.
ISPA OR PSPT0699.
Pseudomonas syringae (pv. tomato).
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=323;
[1]
SEQUENCE FROM N.A.
STRAIN=DC3000;
Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
Berry K., Utterback T., Van Aken S., Feildblyum T., Gwin M.,
Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
Brinkac L., Beanan M., Haft D., Selengut J., Neilson W., Davidsen T.,
White O., Fraser C., Collmer A.;
"Complete sequence of Pseudomonas syringae.";
Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
EMBL; AE016558; AA054241.1; -
TIGR; PSPT0699; -
GO; GO:0016740; F:transferase activity; IEA.
GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
InterPro; IPR000092; Polyprenyl synth.
InterPro; IPR008945; Terpenoid synth.
Pfam; PF03348; polyprenyl synth; 1.
PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
Transferase; Complete proteome.
Q SEQUENCE 295 AA; 31255 MW; 86DA1E6092FB501A CRC64;

Query Match 52.4%; Score 791; DB 16; Length 295;
Best Local Similarity 56.8%; Pred. No. 2e-55;
Matches 167; Conservative 36; Mismatches 91; Indels 0; Gaps 0;

Y 4 LKAYLTVCQERVALDARLPAENILPQTLHQAMRYSVLNGGKRTRELLTYATGQALGLP 63
b 2 IASVQASQSTRVNALEGFQAPSPSLRYEAMRYSVWNGKRVPLLAYAACALGV 51
Y 64 ENVLDPACAVEIHVYSLIHDDLPAMDNDLRRGKPTCHKYDEATAILAGDALQALAF 123
b 62 AEDANGAACAVELIHAYSLVHDDLPAMDNDLRRGKPTCHKYDEATAILAGDALQALAF 121
Y 124 EVLADPGITVDAPARLKMVLTALTRASSGQGVGGQAIIDLSGVGRKLTLPENLHHIHT 183
b 122 TALLAPHLSSRNATRLQWSTLAAAGAGAGWGGQAIIDLSGVGRKLTLPENLHHIHT 181
Y 184 GALIRASVNLAAALSKPDLTCTVAKKLDHYAKCIGLSFQVKDILDIADTATLGTQKGD 243
b 182 GALIRASVNLGALASGQADQARLDALQVYARAVGLAFQVDDILDVESDTATLGTQKGD 241
Y 244 IADKPTYPALLGMACAKQAKQELHEQAVESLTGFGSEADLLRELSLYIERTH 297
b 242 IADKPTYPALLGLDAKGYALELRDQALALSLFDTTAEPLRELYIVERRH 295

RESULT 5
82VD4 PRELIMINARY; PRT; 297 AA.
C Q82VD4
T 01-JUN-2003 (TREMELrel. 24, Created)
T 01-JUN-2003 (TREMELrel. 24, Last sequence update)
T 01-OCT-2003 (TREMELrel. 25, Last annotation update)
E Polyprenyl synthetase (EC 2.5.1.10).
N ISPA OR NE1160.
S Nitrosomonas europaea.
C Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
C Nitrosomonadaceae; Nitrosomonas.
X NCBI_TaxID=915;
N [1]
N SEQUENCE FROM N.A.
C STRAIN=ATCC 19718 / IFO 14298;
C MEDLINE=22586410; PubMed=12700255;
```

```
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773(2003).
DR EMBL; BX321860; CAD85071.1; -
DR GO; GO:0004337; F:geranyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000092; Polyprenyl synth.
DR Pfam; PF00348; polyprenyl synth; 1.
DR PROSITE; PS00723; POLYPRENYL SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 297 AA; 32448 MW; A22BC333E65F3699 CRC64;

Query Match 52.2%; Score 787.5; DB 16; Length 297;
Best Local Similarity 56.9%; Pred. No. 3.9e-55;
Matches 160; Conservative 40; Mismatches 80; Indels 1; Gaps 1;

QY 15 VERALDARLPAENILPQTLHQAMRYSVLNGGKRTRELLTYATGQALGLPENVLDPACAV 74
DB 16 VETCLEKHLPEITNCAPARLDHMYRYSVLGGKRVPLLSFAGELSGADKTHATIAAAV 75
QY 75 EFHIVYSLIHDDLPAMDNDLRRGKPTCHKYDEATAILAGDALQALAFVLANDPQITV 134
DB 76 ELIHVYSLVHDDLPAMDNDLRRGKPTCHKYDEATAILAGDALQALAFVLANDPQITV 134
QY 135 DAPARLKMVLTALTRASSGQGVGGQAIIDLSGVGRKLTLPENLHHIHTGALIRASVNL 194
DB 135 DPHVQLEVMRHLAFAGSRMGAGQAIIDLSGVGRKLTLPENLHHIHTGALIRASVNL 194
QY 195 ALSKPDLTCTVAKKLDHYAKCIGLSFQVKDILDIADTATLGTQKGDIDNKPYPAL 254
DB 195 ARCGRNDETQCKLLDHFACMGALAFQVDDILDVESDTATLGTQKGDIDNKPYPAL 254
QY 255 LGWAGAKQAKQELHEQAVESLTGFGSEADLLRELSLYIER 295
DB 255 LGIKQARELAHELQEQATGVNQFSGEALRQQVDFIVQR 295

RESULT 6
Q9JSMO PRELIMINARY; PRT; 298 AA.
AC Q9JSMO
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Geranyltransferase (EC 2.5.1.10).
GN ISPA OR NMA2226.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
ON [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Z2491 / Serogroup A / Serotype 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagsis K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162758; CAB85437.1; -
DR PIR; E81796; E81796.
DR GO; GO:0004337; F:geranyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
```

[illegible]

```

ID: Q3DEF4      PRELIMINARY;      PRT;      298 AA.
AC Q3DFA4;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Geranylgeranyl pyrophosphate synthase.
GN V10314.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxId=672;
[1]_TaxId=672;
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6."
RL Submitted (DSC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016798; AAC08844.1; -.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000692; Polyprenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyprenyl synth; 1.
DR PROSITE; PS00723; POLYPRENYL_SYNTHET_1; 1.
DR PROSITE; PS00444; POLYPRENYL_SYNTHET_2; 1.
DR Complete proteome.
KW SEQUENCE 298 AA; 32538 MW; 65E59976E668066C CRC64;
Query Match 46.8%; Score 707; DB 16; Length 298;
Best Local Similarity 53.6%; Pred. No. 1.2e-48;
Matches 148; Conservative 45; Mismatches 88; Indels 6; Gaps 2
QY 12 QERVERALDA---RLPAENILPQTLQAMRYSVINGKGRTRPLLTYATGQALGLPENVLID 68
DB 13 QQRNNQNLWLDKLPHQN---QALIQAMRYGLVLAGKE-RPELVATGEMLGCSSQELD 69
QY 69 APACNAVFHYVSLIHDDLPAQNDDILRGKPTCHKYKAYDEATAILAGDALQALAEVLAN 128
DB 70 TPASAEICTHAYSLIHDDLPAQNDDILRGKPTCHKYKAYDEATAILGDTALQTLATLAE 129

```



129 DPGITVDAPARLKMITALTRASGQGMVGGQALDLSVGRKLTLPENLHHKTKGALIR 188  
130 GTLSADGETQVRVMIQALASGAQGCICGQALDLAENRAVLEELERHKTGALIR 189  
189 ASVNLALSKPDLTVCVAKLDHYAKICIGLSFOVKDDILDIEDATTLGKTQCKDINDK 248  
190 CAVRLGALAGEKGLALPOLKYSTAIGLAFVQDDILDIDISDTTLGKPGQSDQELNK 249  
249 PTPALGMAGAKOKAQELHEQAVESLTGFGSEADLLRELSLYIER 295  
250 STYPSLLGLEGMEKHAHTLLHEALQALEAIPYNTQHLERFARVWVER 296

## RESULT 9

Q87RT9 PRELIMINARY; PRT; 291 AA.  
AC Q87RT9; 291 AA.  
DT 01-JUN-2003 (T-EMBLrel. 22, Created)  
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Geranyltransferase.  
GN VP0687.  
OS Xanthomonas axonopodis (pv. citri).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
OC Xanthomonadaceae; Xanthomonas.  
CX NCBI\_TaxID=92829;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=306 / ATCC 13902 / XV 101;  
RX MEDLINE=2022145; PubMed=12024217;  
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,  
A Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,  
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,  
A Camarotte G., Camuana A.M., Cardozo J., Chamberg F., Ciapina L.P.,  
A Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,  
A Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,  
A Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,  
A Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,  
A Locali E.C., Machado M.A., Madeira A.B.N., Martinez-Rossi N.M.,  
A Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,  
A Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
A Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,  
A Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezsa R.I.D.,  
A Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
A Setubal J.C., Kitajima J.P.;  
T "Comparison of the genomes of two Xanthomonas pathogens with differing  
T host specificities."  
L Nature 417:459-463(2002).

EMBL: AB013916; AM37607.1; --  
R GO; GO:0016740; F:transferase activity; IEA.  
R GO; GO:0008299; P:isoprenoid biosynthesis; IEA.  
R InterPro; IPR000092; Polyprenyl synth.  
R InterPro; IPR008949; Terpenoid synth.  
R Pfam; PF00348; polyprenyl synth\_1  
R PROSITE; PS00723; POLYPRENYL\_SYNTHET\_1; 1.  
R PROSITE; PS00444; POLYPRENYL\_SYNTHET\_2; 1.  
R Transferase; Complete proteome.  
Q SEQUENCE 291 AA; 30516 MW; 404C3B5EDFE8BF67 CRC64;

Query Match 46.8%; Score 706; DB 16; Length 291;  
Best Local Similarity 51.7%; Pred. No. 1.4e-48;  
Matches 148; Conservative 44; Mismatches 86; Indels 8; Gaps 3;  
14 RVERDALPRLPAENTLPQTLHOAMRYSVLNGGKTRPLLTATGQALGLPENVLDPACA 73  
12 RTERSLEAGLPSTHAPQRLHAARHVAVLGGKRWKPLLVASGALFGAEDQLDTPAVA 71  
74 VFHFVYSLIHDLPAMDNDLRRGKPTCHKAYDEATLADGALQALAFVNLDPGIT 133  
72 VELIRAYSLVHDDLPMDDDLARGQPTVHAFDDEATLADGALQALAFVNLDPGIT 128  
134 VDAPARLKMITALTRASGQGMVGGQALDLSVGRKLTLPENLHHKTKGALIRASVNL 193

129 ASAEIRVGMQSLATAAGAGMGQALDIDATQLOSLQHLQRMHAKTGTALIRAVRM 188  
194 AALSKPDLTVCVAKLDHYAKICIGLSFOVKDDILDIEDATTLGKTQCKDINDKPT 251  
189 GALTG--GAALADQORLDFADALGAFVQDDILDIDISDTTLGKPGQSDQELNK 245  
252 PALLGMAGAKOKAQELHEQAVESLTGFGSEADLLRELSLYIERTH 297  
246 PALLGMGDGAKLAELAAARMHDVLPVGPQGETLATLARFAVNRAH 291

## RESULT 10

Q87RT9 PRELIMINARY; PRT; 294 AA.  
AC Q87RT9; 294 AA.  
DT 01-JUN-2003 (T-EMBLrel. 24, Created)  
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
DE Geranyltransferase.  
GN VP0687.  
OS Vibrio parahaemolyticus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
CX NCBI\_TaxID=670;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RIND 2210633 / Serotype O3:K6;  
RX MEDLINE=22508454; PubMed=12620739;  
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
RA Iijima Y., Nishima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.,  
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
RT distinct from that of V. cholerae."  
RL Lancet 361:743-749(2003).  
RL EMBL; AP005075; BAC8950.1; --  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.  
DR InterPro; IPR000092; Polyprenyl synth.  
DR InterPro; IPR008949; Terpenoid synth.  
DR Pfam; PF00348; polyprenyl synth\_1.  
DR PROSITE; PS00723; POLYPRENYL\_SYNTHET\_1; 1.  
DR PROSITE; PS00444; POLYPRENYL\_SYNTHET\_2; 1.  
DR Transferase; Complete proteome.  
SQ SEQUENCE 294 AA; 32337 MW; A1C1924A66774FD0 CRC64;

Query Match 46.8%; Score 706; DB 16; Length 294;  
Best Local Similarity 52.1%; Pred. No. 1.4e-48;  
Matches 150; Conservative 41; Mismatches 97; Indels 0; Gaps 0;  
8 LTVQERVERALDARLPAENILPQTLHOAMRYSVLNGGKTRPLLTATGQALGLPENVL 67  
5 LTSFQERNQQLNLWLEQLPYQEQPLIQAMKYGLLGGKRVPLVYITGQMLGCKPEDL 64  
68 DAPACAVEFHVYSLIHDLPAMDNDLRRGKPTCHKAYDEATLADGALQALAFVLA 127  
65 DTPASAEICIHAYSLIHDLPAMDNDLRRGQPTCHIRFDEATLUTGALQTLAFTIIA 124  
128 NDPGITVDAPARLKMITALTRASGQGMVGGQALDLSVGRKLTLPENLHHKTKGALI 187  
125 DGPLSPSAENQVRNMLKALHSSGANGCMVQALDGAENQVSLAEHIEHKTGTGALI 184  
188 RASVNLALSKPDLTVCVAKLDHYAKICIGLSFOVKDDILDIEDATTLGKTQCKDIND 247  
185 DCVAVKLGAALAGEKGLVPLHLYRYSKAIGLAFVQDDILDIDISDTTLGKPGQSDQELN 244  
248 KPTYPALLGMAGAKOKAQELHEQAVESLTGFGSEADLLRELSLYIER 295  
245 KSTYPSLLGLEGMEKHAHTLLHEALQALEAIPYNTQHLERFARVWVER 292

## RESULT 11

Q8ZED0



```
X NCBI_TaxID=666;
P [1]
N SEQUENCE FROM N.A.
C STRAIN=El Tor N16961 / Serotype O1;
C MEDLINE=20406833; PubMed=10952301;
X Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
A Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
A Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
A Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
A McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
A Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
A Fraser C.M.;
T "DNA sequence of both chromosomes of the cholera pathogen Vibrio
T cholerae.";
L Nature 406:477-483 (2000).
L R EMBL; AE004173; AAF94052.1; -.
R PIR; A82267; A82267.
R TIGR; V00890; -.
R GO; GO:0016740; F:transferase activity; IEA.
R GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
R InterPro; IPR000892; Polyrenyl synth.
R Pfam; PF00348; polyrenyl synth.
R PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
R PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
W Transferase; Complete proteome.
Q SEQUENCE 294 AA; 32259 MW; E7D33C0FD935CB96 CRC64;
```

```
Query Match 46.6%; Score 704; DB 16; Length 294;
Best Local Similarity 51.2%; Pred. No. 2e-48;
Matches 149; Conservative 45; Mismatches 91; Indels 6; Gaps 2;
```

```
8 LTVQQRVERALDA---RLPAENILPQTILQAMRYSVLNGKRTPLITYATGALGPE 64
: : : : : : : : : : : : : : : : : : : : : : : : : :
5 LSSYQQRNNQDMLNRPFTQL---PLIEAMRYGLLGGKRRPPIVYITGQMLGCEL 61
: : : : : : : : : : : : : : : : : : : : : : : : : :
65 NVLDAPACAVEFHVYSLIHDDLPAWDDNLLRGKPTCHKAYDEATALLAGDALQALFE 124
: : : : : : : : : : : : : : : : : : : : : : : : : :
62 SLDLTPASAVECHAYSLIHDDLPAWDDNLLRGKPTCHKAYDEATALLAGDALQALFE 121
: : : : : : : : : : : : : : : : : : : : : : : : : :
125 VLANDPGITVDAPARKMITALTTRASGQGMVGGQAI DLGVSGRKLTLPLENMHIHRTG 184
: : : : : : : : : : : : : : : : : : : : : : : : : :
122 ILAEGSLAAGETQRYVAMQLAAEASAGQGMCLGQALDLAAENRILSLELETHRNKTG 181
: : : : : : : : : : : : : : : : : : : : : : : : : :
185 ALIRASVNLAAASKPDLDTCAVKLDHYAKTGLSFGVKKDDILDIEDATATLGTGKQDI 244
: : : : : : : : : : : : : : : : : : : : : : : : : :
182 ALMRCALRIGLAAAGKGRAMPDPLDRYAEAVGLAFQVQDDILDIISDTETLGRPQSDQ 241
: : : : : : : : : : : : : : : : : : : : : : : : : :
245 DNDKPTYPALLGAGAKAKOAEHQAEVSLTGFSGEADLLRELSLYIERTH 295
: : : : : : : : : : : : : : : : : : : : : : : : : :
242 ELNKSTYPALLGLEGAQQAHTLLQBALLAEAIPTNTEHLEEFARYVVER 292
: : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 14
82C46
D Q8ZC46 PRELIMINARY; PRT; 306 AA.
C Q8ZC46;
T 01-MAR-2002 (TRENBLrel. 20, Created)
T 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
I 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
E Geranyltransferase (EC 2.5.1.10).
S ISPA OR YP03176 OR Y1009.
S Versinia pestis.
S Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
S Enterobacteriaceae; Versinia.
X NCBI_TaxID=632;
Y [1]
Y SEQUENCE FROM N.A.
P STRAIN=CO-92 / Biovar Mediaevalis;
P MEDLINE=21470413; PubMed=11586360;
P Parkhill J., Wren B.W., Thomson N.R., Fitball R.W., Holden M.T.G.,
P Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
P Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
```

```
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Kariyshev A.V.,
RA Leather S., Moulé S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds W., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Versinia pestis, the causative agent of plague."
RL Nature 413:523-527 (2001).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=KIMS / Biovar Mediaevalis;
RP MEDLINE=22137863; PubMed=12142430;
RX Deng W., Burland V., Plunkett G., Boutin A., Mayhew G.F., Liss P.,
RX Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Versinia pestis KIM."
RL J. Bacteriol. 184:4601-4611 (2002).
DR EMBL; AJ414155; CAC92411.1; -.
DR EMBL; AE013704; AAM84590.1; -.
DR PIR; AH0385; AH0385.
DR GO; GO:0004337; F:geranyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008299; P:isoprenoid biosynthesis; IEA.
DR InterPro; IPR000892; Polyrenyl synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00348; polyrenyl synth.
DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.
DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 306 AA; 32919 MW; 84F61276DFB6E106 CRC64;
```

```
Query Match 46.5%; Score 702.5; DB 16; Length 306;
Best Local Similarity 52.5%; Pred. No. 2.8e-48;
Matches 157; Conservative 38; Mismatches 95; Indels 9; Gaps 3;
```

```
5 KAYLVQQRVERAL---DARLPAENILPQTILQAMRYSVLNGKRTPLITYATGQALG 61
: : : : : : : : : : : : : : : : : : : : : : : : : :
11 KQOLAHQQRVNCALLDFTIAPLPGN---SNLIEAMRYCAVIGGKRLPYLYATGQMGF 67
: : : : : : : : : : : : : : : : : : : : : : : : : :
62 LPENVLDAPACAVEFHVYSLIHDDLPAWDDNLLRGKPTCHKAYDEATALLAGDALQAL 121
: : : : : : : : : : : : : : : : : : : : : : : : : :
68 LSLANDAPAAIECHAYSLIHDDLPAWDDNLLRGKPTCHKAYDEATALLAGDALQAL 127
: : : : : : : : : : : : : : : : : : : : : : : : : :
122 APEVLANDPGITVDAPARKMITALTTRASGQGMVGGQAI DLGVSGRKLTLPLENMHIH 181
: : : : : : : : : : : : : : : : : : : : : : : : : :
128 AFSILAEAMPVADKXDRLSMIIELAAQASGAGMCAGALDLAEATRPVLLGDLEQIHRH 187
: : : : : : : : : : : : : : : : : : : : : : : : : :
182 KTGALIRASVNLAAASKPDLDTCAVKLDHYAKTGLSFGVKKDDILDIEDATATLGTGKQ 241
: : : : : : : : : : : : : : : : : : : : : : : : : :
188 KTGALIRAAVRLGAQAAGEAGERTLLSLDRYAEATGLAFQVQDDILDVIGTATIGKRG 247
: : : : : : : : : : : : : : : : : : : : : : : : : :
242 KDIDNDKPTYPALLGAGAKAKOAEHQAEVSLTGFSGEADLLRELSLYIERTH 297
: : : : : : : : : : : : : : : : : : : : : : : : : :
248 SDQQLGKSTYPALLGLDCAQTFRMDLYQAEALDELAKQSNITDLSQALARFIIRNN 306
: : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 15
Q8XE75
ID Q8XE75 PRELIMINARY; PRT; 299 AA.
AC Q8XE75;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Geranyltransferase (farnesyl diphosphate synthase).
GN ISPA OR Z0524 OR ECS0475.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RC MEDLINE=21074935; PubMed=11206551;
```

Search completed: February 29, 2004, 14:50:36  
Job time : 31.4213 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: February 29, 2004, 14:34:14 ; Search time 8.26685 Seconds

(without alignments)  
3455.835 Million cell updates/sec

US-09-941-947A-20

title:

effect score: 1510

sequence: 1 MSKLKAVLTVCQVERALD.....FGSEADLLRELSLYIERTH 297

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 283366 seqs, 96191526 residues

real number of hits satisfying chosen parameters: 283366

imum DB seq length: 0

aximum DB seq length: 20000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

PIR 78:\*\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	831	55.0	295	2 F83139	Geranyltranstransf
2	729	48.3	298	2 E81796	Geranyltranstransf
3	705.5	46.7	299	2 AG0554	Geranyltranstransf
4	704	46.6	294	2 A82267	Geranyltranstransf
5	702.5	46.5	306	2 AH0385	Geranyltranstransf
6	695.5	46.1	299	2 C90688	Geranyltranstransf
7	695.5	46.1	299	2 G85538	Geranyltranstransf
8	691.5	45.8	299	2 JQ0665	Geranyltranstransf
9	677	44.7	291	2 D82778	Geranyltranstransf
10	675	44.7	291	2 F81217	Geranyltranstransf
11	668	44.2	259	2 S74538	Geranylgeranyl pyr
12	606.5	40.2	302	2 F83434	Geranylgeranyl pyr
13	604	40.0	371	2 F53722	farnesyltranstransf
14	596	39.5	369	2 S53722	farnesyltranstransf
15	590.5	39.1	366	2 T10452	farnesyltranstransf
16	586	38.8	309	2 AE1833	Geranylgeranyl dip
17	583	38.6	332	2 T10875	y4ku protein - Rhi
18	578	38.3	332	2 I40213	probable dimethyla
19	577	38.2	297	2 JX0257	Geranyltranstransf
20	575	38.1	357	2 T09966	farnesyltranstransf
21	567.5	37.6	347	2 E84566	probable geranylge
22	565	37.4	294	2 E83997	Geranyltranstransf
23	565	37.4	372	2 G84566	probable geranylge
24	557.5	36.9	304	2 A13285	Geranyltranstransf
25	552.5	36.6	316	2 T11021	farnesyltranstransf
26	541.5	35.9	294	2 D87505	Geranyltranstransf
27	537	35.6	293	2 AC1245	Geranyltranstransf
28	531	35.2	272	2 A69961	Geranyltranstransf
29	531	35.2	376	2 T02429	Geranylgeranyl pyr

## RESULT 2

E81796

Geranyltranstransferase (EC 2.5.1.10) NMA2226 [imported] - Neisseria meningitidis (strai

C;Species: Neisseria meningitidis

C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C;Accession: E81796

30 529 35.0 293 2 AG1607 geranyltranstransf  
31 522.5 34.6 300 2 T06969 farnesyltranstransf  
32 522 34.6 378 2 S71230 geranylgeranyl pyr  
33 517.5 34.3 293 2 A89932 hypothetical prote  
34 517 34.2 304 2 F97685 geranyltranstransf  
35 517 34.2 335 2 AH2910 geranyltranstransf  
36 501 33.2 282 2 B84984 geranyltranstransf  
37 488.5 32.4 285 2 A86732 geranyltranstransf  
38 476.5 31.6 289 2 D97156 probable geranylge  
39 446.5 29.6 291 2 G95139 geranyltranstransf  
40 445.5 29.5 291 2 F98007 geranyltranstransf  
41 440.5 29.2 222 2 S71231 geranylgeranyl pyr  
42 399 26.4 289 2 S04407 phytoene synthase  
43 391 25.9 303 2 A64636 geranyltranstransf  
44 379 25.1 281 2 B81261 geranyltranstransf  
45 375 24.8 303 2 G71878 geranyltranstransf

## ALIGNMENTS

### RESULT 1

F83139

Geranyltranstransferase PA4043 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C;Accession: F83139

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bz  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: F83139

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-295 x850>

A;Cross-references: GB:AE004821; GB:AE004091; NID:g9950236; PIDN:AAG07430.1; GSPDB:GN001

A;Experimental source: strain PA01

C;Genetics:

A;Gene: is; PA4043

Query Match 55.0%; Score 831; DB 2; Length 295;

Best Local Similarity 60.3%; Pred. No. 2.6e-58;

Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;

Qy 4 LKAYLTVCQVERALDARLPAENILPOTLHOAMRYSVLNGGKTRPLLTATGQALGLP 63

Db 2 IAAQACQARVDALDALFVAPRELQRLYEAMRYSVNGGKRVPLLAYACEALGGA 61

Qy 64 ENVLDPACAVFPIVYSLIHDDLPAMDNDLRRKPTCHKAYDEATAILAGDALQALAP 123

Db 62 PORADAAACAVELIHAYSLVHDDLPAMDNDLRRKQPTTHAFDEATAILAADGLQALAP 121

Qy 124 EVLANDPGITVDAPARLKMITALTRASSQGVGGQALDLSGVGSKLTLPLENNHIHKT 183

Db 122 EVLADTRNPQHVAVCLEMLTILARRAGSAGVGGQALDLSGVGVALDQAALEVWHRKT 181

Qy 184 GALIRASVNLAAALSKPDLDTVCVAKKLDEHYAKTIGHSFQVXDDILDIEDATATLTKGTQKD 243

Db 182 GALIEASVRLGALAGRAEPASIAALRYEAYEAIGLAFQVQDDILDVESDTATLTKGTQKD 241

Qy 244 IONDKPTYPALIGMAGAKQKQELHEQAVESLTGFGSEADLLRELSLYIER 295

Db 242 QAHNKPTYPALIGLEAAKGYALELRDLAALDGPSPADPLRLQARYIVER 293

62	STATUDAPAAAVECIIHAYSIHDDLPAMUDDDDURRGUPTCHIKFGBANALAGDALOTUA	121
123	FEVLNAPGIIITVDAPARKOMITALTTRASSQGMVGQOALDLSGVBGKLTLPBLENMHTK	182
122	FTIISDAPMPEVADORDIAMIAELANASGIAGMCGQOALDLAEGORITLDALEIRIHRK	181
103	TGALIRASVNI.AALS.KPD.LOTCVAKKL.DHYKACIGI.SF.VKDDILDIEADTATL.GKNTQK	242
182	TGALIRAAVR.LGALSAGD.KGENTLPILDRVAES.IGLAFVQDDILDVVGDTATL.SKRQQA	241
243	DI.DNDKPTYPALLGMACAKQKQELHEQAVESJTGFGSEA---DILRELSLVIIER	295
242	DOOLCKSTYPALLGLEARNKARDLTEDAROS.SHO.LAAOSLPTSAL.EALNTIIOR	297

RESULT 4  
A82267 geranyltransferase VC0890 [imported] - Vibrio cholerae (strain N16961 serotype O1)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-20001  
C:Accession: A82267  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.; DiMarco, A.;  
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.;  
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae  
A:Reference number: A82035; MUID:20406833; PMID:10952391  
A:Accession: A82267  
A:Status: preliminary  
A:Molecule type: DNA  
A:ResIdues: 1-294 <HEI>  
A:Cross-references: GB:AE004173; GB:AE003852; NID:G9555341; PIDN:AAF94052.1; G:  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC0890  
A:Map position: 1  
C:Subfamily: dimethylallyltransferase

	Best Local Similarity	51.2%; Piled. NO. 2.25e+46;	Mismatches 45;	Indels 91;	Gaps 6;
QY	8	LTVQOEVERALDA--RLPAENILPQTLHQAMRYSVLWNGKRETRPLLTATGQALGLPE	64		
Db	5	LSSVQQRNNQQLDQWLNRIPQTL---PLTEAMRYGLLLGCKREARPLYVITQGLMGCEL	61		
QY	65	NVLDPACAFEFTHVYSLIHDDLPAMDNDLREGKCTCHKAYDEATAILLAGDALQALAE	124		
Db	62	SLDLPASAVECHAYSLIHDDLPAMDDELREGKCTCHIQDEATAITGDAQLQIATF	121		
QY	125	VLANDPGITVDPAARLKMVITALTASGQGVGGQAIILGSGVRKLTLPLELNMMHKTG	184		
Db	122	ILAEGLLSAAGETQVRVAMLQALAEASGAQGVCLGQALDLAAENRLISLELETIHNKGT	181		
QY	185	ALIRASVNLAAALSKPDLDTCVAKLIDHYAKICGLSFQVKKDILDLTDADATLTKTGKDI	244		
Db	182	ALMCAKRLGALAGEKGRAMLPELQRYAEAVGLAFQVQDDILITISDTETLTKPGQSDQ	241		
QY	245	DNDKPTPALGEMAGAKQKQAEELFEQAVESLTGFGSEADLELSLYITIER	295		
Db	242	EINKSTPTPALGLEGAAKQKHTLLORALLALAEALPYNTEHLEEFARYVWER	292		

RESULT 5  
AH0385  
geranyltranstransferase [EC 2.5.1.10] [imported] - Yersinia pestis (strain CO92)  
C.Species: Yersinia pestis  
C.Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 17-May-2002  
C.Accession: AHQ385  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice,  
deno-Farrago, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Douc  
ill, M.; Rutherford  
S.; Simmonds, K.; Stevens, K.; Skelton, J.; Whitehead, S.;

```
ature 413, 523-527, 2001
;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
;Reference number: AB0001; MUID:21470413; PMID:11586360
;Accession: AH0385
;Status: preliminary
;Molecule type: DNA
;Residues: 1-306 <KUR>
;Cross-references: GB:AL590842; PIDN:CAC92411.1; PID:g15981114; GSPDB:GN00175
;Genetics:
;Gene: ispa
;Superfamily: dimethylallyltranstransferase
;Keywords: transferase

Query Match 46.1%; Score 702.5; DB 2; Length 306;
Best Local Similarity 52.4%; Pred. No. 1.4e-47;
Matches 157; Conservative 38; Mismatches 95; Indels 9; Gaps 3;

y 5 KAVLTWCQERVERAL---DARLPAENILPQTLHQAMRYSLVNGGKTRPLLTATGQALG 61
b 11 KQGLAAHQQRVNOALDFIAPLPFGN---SNLEAMRYGAVTGGKRLRPVLYATGQWFG 67
y 62 LPENVLDAPACAVEFHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQAL 121
b 68 LSLANLDAFAAECHAYSLIHDDLPAMDNDLRRGKPTCHVKFGEAHAILAGDALQTL 127
y 122 AFEVLANDPGITVDAPARLKMITALTRASSGQMGVGGQIDLGSGVRKLTLPLELNMHIH 181
b 128 AFSILAEAPMPDVADKRLSMITELAQASGAACMCAGQALDLEASTRPVLLGDLEQIHRH 187
y 182 KTCALIRASVNLAAALSKPDLDTCAKLDHYAKCIGLSFOVKDDILDIDEADTATLGKTOG 241
b 188 KTCALIRAAVRLGAQAGAGRGKTLISLDRYAEAGLAFQVQDDILDVIGDTATIGKQOG 247
y 242 KTDNDKPTYPALLGWAGAKQAQELHEQAVESLTGFGSEA---DLLRELSLYIERTH 297
b 248 SDQQLGKSTYPALLGLDCAQTRAWDLYQEALEBALDELAQSYNTDSLOALAFRIERN 306

RESULT 6
90688
eranyltranstransferase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05
;Species: Escherichia coli
;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 17-May-2002
;Accession: C90688
;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
;asawara, N.; Yasunaga, T.; Kunara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
;Reference number: A99629; MUID:21156231; PMID:11258796
;Accession: C90688
;Status: preliminary
;Molecule type: DNA
;Residues: 1-299 <HAY>
;Cross-references: GB:BA000007; PIDN:BA033898.1; PID:g13359932; GSPDB:GN00154
;Experimental source: strain O157:H7, substrain RIMD 0509952
;Genetics:
;Gene: ECs0475
;Superfamily: dimethylallyltranstransferase

Query Match 46.1%; Score 695.5; DB 2; Length 299;
Best Local Similarity 52.4%; Pred. No. 1.4e-47;
Matches 154; Conservative 35; Mismatches 96; Indels 9; Gaps 3;

y 8 LTVCOERVERALD---ARLPAENILPQTLHQAMRYSLVNGGKTRPLLTATGQALGLPE 64
b 7 LEACVQANQALSRRFIAPLPFQN---TPVETWYQALGKRLRPVLYATGEMFGVST 63
y 65 NVLDAPACAVEFHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFE 124
b 64 NTLADAPAAAVECHYSLIHDDLPAMDNDLRRGKPTCHVKFGEANAILAGDALQALAFS 123
y 125 VLANDPGITVDAPARLKMITALTRASSGQMGVGGQIDLGSGVRKLTLPLELNMHIHKTG 184
b 124 IESDADMPESVDRDRISMISELASGIAGMCGGQALDLDAGKGVPLDALARIHRTG 183
y 185 ALIRASVNLAAALSKPDLDTCAKLDHYAKCIGLSFOVKDDILDIDEADTATLGKTOGKDI 244
b 184 ALIRAAVRLGALSAGDKGRRALPVLDDKYAESIGLAFQVQDDILDVVGDTATLGKRCQADQ 243
y 245 DNDKPTYPALLGWAGAKQAQELHEQAVESLTGFGSEA---DLLRELSLYIERT 295
b 244 QLCKSTYPALLGLEQARKKAQDLIDARQSLAQARQSIDTSALEALADYIIQR 297

RESULT 8
JQ0665
Geranyltranstransferase (EC 2.5.1.10) - Escherichia coli (strain K-12)
;Species: Escherichia coli
;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 01-Mar-2002
;Accession: JQ0665; E64771
;R; Fujisaki, S.; Hara, H.; Nishimura, Y.; Horiuchi, K.; Nishino, T.
;J. Biochem. 108, 995-1000, 1990
;Title: Cloning and nucleotide sequence of the ispa gene responsible for farnesyl diphos
;Reference number: JQ0664; MUID:91210228; PMID:2089044
;Accession: JQ0665
;Molecule type: DNA
;Residues: 1-299 <FUJ>
;Cross-references: GB:D00694; NID:g216582; PIDN:BA00539.1; PID:g216584
;Experimental source: strain K-12
;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
```

A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A>Title: The complete genome sequence of *Escherichia coli* K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: E64771  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-299 <BLAT>  
A;Cross-references: GB:AB00148; GB:U00096; NID:g1786614; PIDN:AAC73524.1; PID:g1786623;  
A;Experimental source: strain K-12, substrain MG1655  
A;Comment: This enzyme catalyzes the condensation of isopentenyl diphosphate with dimethyl  
3;Genetics:  
A;Gene: ispA  
A;Map position: 10 min  
A;Function:  
A;Superfamily: isoprenoid biosynthesis  
A;Pathway: isoprenoid biosynthesis  
A;Keywords: ATP; isoprenoid biosynthesis; nucleotide binding; P-loop; transferase  
F;84-93/Region: aspartate-rich  
F;223-227/Region: DDXD motif  
F;1241-248/Region: nucleotide-binding motif A (P-loop)

Query Match 45.8%; Score 691.5; DB 2; Length 299;  
Best Local Similarity 52.0%; Pred. No. 2.8e-47;  
Matches 153; Conservative 36; Mismatches 96; Indels 9; Gaps 3;

2Y 8 LTVQCOERVALD---ARLPAENILPOTLHOAMRYSVINGCKTRPLLYATQALGLPE 64  
2b 7 LEACVQKQANQAFIAPLFPQN---TPVETWQYGAAGKRLRPPLVYATGMPGVST 63  
2Y 65 NVLDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 124  
2b 64 NTLDPAAAVECHAYSLIHDDLPAMDNDLRRGLPCHVKGKGNAILAGDALQALAFS 123  
2Y 125 VLNDPGITVDAPARLKMITALTFRASQSGMVGQAIDGSGVRKKTLPLENNHIHTG 184  
2b 124 ILSDAMPVSDRDRISMISLASAGTAGCGQALDLDAEGKVPFLDALEIRHRTG 183  
2Y 185 ALIRASVNLALSKPDLTCAKLDHYAKCIGLSFOVKDDILDEADTATLCTGKDI 244  
2b 184 ALIRAVRLGALSGDKRRALPVLKYASIGLAGFVQDDILDVGDATLIGRGQADQ 243  
2Y 245 DNDKPTYPALLGMAGAKQKQAEHQAVESLTFGSGSA---DLLRELSYIIR 295  
2b 244 QLGRSTYPALLGCAKQKARDLDDARQSLKOLAEQSLDTSALEADYIIOR 297

RESULT 9  
364123  
geranyltransferase (EC 2.5.1.10) - Haemophilus influenzae (strain Rd KX20)  
A;Species: Haemophilus influenzae  
A;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 29-Sep-1999  
A;Accession: C64123  
A;Authors: R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirnesh, E.P.; Kerlavage, A.  
Gocayne, J.D.; Scott, J.; Shirley, R.; Lib, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:95350630; PMID:7542800  
A;Accession: C64123  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-295 <TTGR>  
A;Cross-references: GB:U32822; GB:L42023; NID:g1574265; PIDN:AAC23087.1; PID:g1574277; T  
A;Genetics:  
A;Gene: ispA  
A;Function:  
A;Description: catalyzes condensation of isopentenyl diphosphate with dimethylallyl diph  
A;Pathway: isoprenoid biosynthesis  
A;Superfamily: dimethylallyltransferase  
A;Keywords: isoprenoid biosynthesis; transferase  
A;85-98/Region: aspartate-rich

F;223-227/Region: DDXD motif

Query Match 44.8%; Score 677; DB 2; Length 295;  
Best Local Similarity 49.0%; Pred. No. 3.9e-46;  
Matches 145; Conservative 50; Mismatches 97; Indels 4; Gaps 3;

QY 1 MSKLKAVITVQCOERVALDARLPAENILPOTLHOAMRYSVINGCKTRPLLYATQAL 60  
Db 1 MGHFSEELQVQVTEINRFLQAQFEGIESHNAPLEAMKYALLGKRVKRVPLVYATQOML 60  
QY 61 GLPENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQA 120  
Db 61 GAEKQTLDYAAAIEATHAYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQS 120  
QY 121 LAPEVLNDPGITVDAPARLKMITALTFRASQSGMVGQAIDGSGVRKKTLPLENNHI 180  
Db 121 FAFELITKTPNISTE--QKLAIIQILAQAGAGVQCMCLGSLDLTISEHKQISLSLEIHR 178  
QY 181 HKTGALIRASVNLALSKPDL--DTCVAKLDHYAKCIGLSFOVKDDILDEADTATLTKT 239  
Db 179 NKTGALLIATLALKGFCISPHPTDKRLQESITQYAAIGLAGFVQDDILDEGDSASIGKQ 238  
QY 240 QGKDIDNDKPTYPALLGMAGAKQKQAEHQAVESLTFGSGSEADLRELSYIIR 295  
Db 239 VGADLDLDKSTYKPLGLSGAKQKQADLYQSALSLEKIPEDT-TVRALAEFIIR 293

RESULT 10  
D82778  
geranyltransferase (farnesyl-diphosphate synthase) XF0661 [imported] - Xylella fas  
A;Species: Xylella fastidiosa  
A;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 17-Nov-2000  
A;Accession: D82778  
A;Anonymous: The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ  
Nucleotide 406, 151-157, 2000  
A;Title: The genome sequence of the plant pathogen Xylella  
A;Reference number: A82515; MUID:20365717; PMID:10910347  
A;Note: for a complete list of authors see reference number A59328 below  
A;Accession: D82778  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-291 <SIM>  
A;Cross-references: GB:AE003910; GB:AE003849; NID:g9105532; PIDN:AAF83471.1; GSPDB:GN00  
A;Experimental source: strain 9a5c  
A;Authors: A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, J.  
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, J  
A;Authors: Martins, E.M.P.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.;  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasal  
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshakho, M.H.; Vallada, H.; van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;  
A;Reference number: A59328  
A;Contents: annotation  
A;Gene: XF0661  
A;Superfamily: dimethylallyltransferase

Query Match 44.7%; Score 675; DB 2; Length 291;  
Best Local Similarity 50.2%; Pred. No. 5.6e-46;  
Matches 149; Conservative 37; Mismatches 99; Indels 12; Gaps 3;

QY 5 KAVITVQCOERVALDARLPAENILPOTLHOAMRYSVINGCKTRPLLYATQALGLPE 64  
Db 3 BALTRWRQRTDSVLETLISPTQAPQRLHAAMRYVLSKGRILPLVYAGHLFVDS 62  
QY 65 NVLDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 124  
Db 63 PLLDVPAAAVELIHAYSLVHDDLPAMDNDLRRGRPTVTHAFDEATAILGDLQTLAF 122



A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. S.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S74538

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-302 <KAN>

A:Cross-references: EMBL:D90899; GB:AB001339; NID:G1651650; PIDN:BAAL6690.1; PID:G1651707

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: Geranyltransferase

Query Match 40.2%; Score 606.5; DB 2; Length 302;

Best Local Similarity 47.2%; Pred. No. 1.5e-40;

Matches 143; Conservative 42; Mismatches 97; Indels 21; Gaps 6;

QY 4 LKAVITVCOQVERALDARLPAENILPOTLHQAMRYSVLNGKTRPLLTATGALGLP 63

DB 12 LAQVLOVKGVVEAALDSSLAIR--PEKIVEAMRYSLLAGKRLRPLILCITACELCGD 69

QY 64 ENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 123

DB 70 EALALPACALEMHTVSLIHDDLPAMDNDLFRGKPTCHKAYDEATAILAGDALQALAF 129

QY 124 E-VLANDPGITVDAPARKMITALTTRASSQGVGGQALDGLSVGRKLTPLPE-LENMHIH 181

DB 130 EYVTHTP--QADPQALLQVIALRGRTVGAAGLVGGQVLDLSEGRDITPETLTPIHTH 187

QY 182 KTGALIRASVNLALSKPDLDTCA-----KKLDHYAKICGLSVKDDIILDIADTA 234

DB 188 KTGALLERASVLTGAI-----LAGATGQQQRRLARIQNIGLAFQVDDIILDTATQE 239

QY 235 TLGKTQKGDINDKPTYPALLGMAGAKQAEHLHQAVESLVTGSSADLRLRLSLYIE 294

DB 240 ELGKTACKQVKAQKATYPSLLGLEASRAQSLIDQAIVALESPGSAEPQALAEVIVA 299

QY 295 RTH 297

DB 300 RKY 302

RESULT 13

F85434

geranylgeranyl pyrophosphate synthase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 24-Aug-2001

C:Accession: F85434

C:Anonymous: The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory

A:Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.

A:Reference number: AB5001; MUID:20083488; PMID:10617193

A:Accession: F85434

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-371 <STO>

A:Cross-references: GB:NC\_001268; NID:G7270630; PIDN:C2B80347.1; GSPDB:GN00140

C:Genetics: ATg36810

A:Map position: 4

C:Superfamily: geranyltransferase

Query Match 40.0%; Score 604; DB 2; Length 371;

Best Local Similarity 45.9%; Pred. No. 3.1e-40;

Matches 136; Conservative 51; Mismatches 99; Indels 10; Gaps 6;

QY 6 AYLTVCQVERALDARLPAENILPOTLHQAMRYSVLNGKTRPLLTATGALGLP 65

DB 78 SYITTKAELVNKALDGAVALP--PEKIVEAMRYSLLAGKRLRPLILCITACELCGD 135

QY 66 VLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 125

DB 136 TAMPAACAVEMHTVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 195

QY 126 LANDPGITVDAPAR-LKMITALTTRASSQGVGGQALDGLSVGRKLTPLPELENMHIH 191

2Y 125 VLANDPGITVDAPARLKMITALTTRASSQGVGGQALDGLSVGRKLTPLPELENMHIHTG 184

2b 123 CLADAPS---DTSLEIAWLOTLTAAGVAGCGGQAIDIDATGSGQTLKALEMHALKSG 179

2Y 185 ALIRASVNLAL-----SKPDLDTCAVKKLHYAKICGLSVKDDIILDIADTATLGKTQ 240

2b 180 TLMRAAVTMGALAGDPSAEDLH-----LTFPATLGLAQVDDIILDISSSEQLOKTA 234

2Y 241 GKQDINDKPTYPALLGMAGAKQAEHLHQAVESLVTGSSADLRLRLSLYIERTH 297

2b 235 GKDAIQKSTPALLSVEGAKCYLQELAEALYTLQHPYGERAAPLTALARLAVERAH 291

RESULT 11

F81217

geranyltransferase NMB0261 [imported] - Neisseria meningitidis (strain MC58 serogroup B)

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C:Accession: F81217

C:Tettein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; H. Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzia, M.

C:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.

C:Reference number: AB1090; MUID:20175755; PMID:10710307

C:Accession: F81217

C:Status: Preliminary

C:Molecule type: DNA

C:Residues: 1-259 <TET>

C:Cross-references: GB:AB002098; GB:AB002383; NID:G7225484; PIDN:AAAF0715.1; PID:G722548

C:Experimental source: serogroup B, strain MC58

C:Genetics: NMB0261

C:Superfamily: dimethylallyltransferase

Query Match 44.2%; Score 668; DB 2; Length 259;

Best Local Similarity 52.5%; Pred. No. 1.7e-45;

Matches 137; Conservative 39; Mismatches 83; Indels 2; Gaps 1;

Y 37 MRYSLVNGKTRPLLTATGALGLPENVLDPACAVEFIHVYSLIHDDLPAMDNDL 96

b 1 MRYAALDGGKRLPMLVLAASELGEAVHEAQVQAEMIHVYSLVHDDMPAMDNDL 60

Y 97 RGKPTCHKAYDEATAILAGDALQALAFVLANDPGITVDAPARKMITALTTRASSQGV 156

b 61 RGKPTCHKYGEATALTGALQTPDVLSPRE--TELPAARQLAMSLVLRAGSGRGA 118

Y 157 GGQAIDGLSVGRKLTPLPELENMHIHTGALIRASVNLALSKPDLDTCAVKKLHYAKI 216

b 119 GGQAIDLANVQKQVADLQEMSLKTGALIRAAVLLGATACPDLSDBLSVLDIYAANK 178

Y 217 GLSPQVKKDILDTADTATLGKTQKGDINDKPTYPALLGMAGAKQAEHLHQAVESLT 276

b 179 GLAQVDDVLDCEADTATLGKTAGKQANDKPTTYKMLGLEARSYAKLVLAERVALE 238

Y 277 GFQSEADLRLSLYIERTH 297

b 239 PFQDKALRLQLAEFAVARKY 259

RESULT 12

74538

geranylgeranyl pyrophosphate synthase - *Synechocystis* sp. (strain PCC 6803)

C:Alternate names: hypothetical protein slr0739

C:Species: *Synechocystis* sp.

C:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: S74538

C:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; K. Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

C:Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. S.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S74538

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-302 <KAN>

A:Cross-references: EMBL:D90899; GB:AB001339; NID:G1651650; PIDN:BAAL6690.1; PID:G1651707

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Superfamily: Geranyltransferase

Query Match 40.2%; Score 606.5; DB 2; Length 302;

Best Local Similarity 47.2%; Pred. No. 1.5e-40;

Matches 143; Conservative 42; Mismatches 97; Indels 21; Gaps 6;

QY 4 LKAVITVCOQVERDALDARLPAENILPOTLHQAMRYSVLNGKTRPLLTATGALGLP 63

DB 12 LAQVQVKGVVEAALDSSLAIAAR--PEKIVEAMRYSLLAGKRLRPLILCITACELCGD 69

QY 64 ENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 123

DB 70 EALALPACALEMHTVSLIHDDLPAMDNDLFRGKPTCHKAYDEATAILAGDALQALAF 129

QY 124 E-VLANDPGITVDAPARKMITALTTRASSQGVGGQALDGLSVGRKLTPLPE-LENNMHIH 181

DB 130 EYVTHTP--QADPQALLQVIALRGRTVGAAGLVGGQVLDLSEGRDITPETLTPIHTH 187

QY 182 KTGALIRASVNLALSKPDLDTCA-----KKLDHYAKICGLSVFQKDDILDIADTA 234

DB 188 KTGALLERASVLTGAI-----LAGATGQQQRRLARIQNIGLAFQVDDIILDTATQE 239

QY 235 TLGKTQKGDINDKPTYPALLGMAGAKQAEHLHQAVESLITGFGSRADLRLSLYIIE 294

DB 240 ELGKTACKQVKAQKATYPSLLGLEASRAQSLIDQAIVALESPGSAEPQALAEVIVA 299

QY 295 RTH 297

DB 300 RKY 302

RESULT 13

F85434

geranylgeranyl pyrophosphate synthase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 24-Aug-2001

C:Accession: F85434

C:Anonymous: The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory

A:Title: Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*.

A:Reference number: AB5001; MUID:20083488; PMID:10617193

A:Accession: F85434

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-371 <STO>

A:Cross-references: GB:NC\_001268; NID:G7270630; PIDN:C2B80347.1; GSPDB:GN00140

C:Genetics: ATg36810

A:Map position: 4

C:Superfamily: geranyltransferase

Query Match 40.0%; Score 604; DB 2; Length 371;

Best Local Similarity 45.9%; Pred. No. 3.1e-40;

Matches 136; Conservative 51; Mismatches 99; Indels 10; Gaps 6;

QY 6 AYLTVCQVERDALDARLPAENILPOTLHQAMRYSVLNGKTRPLLTATGALGLPEN 65

DB 78 SYITTKAELVNKALDSAVPLRE--PEKIVEAMRYSLLAGKRLRPLILCITACELVGGES 135

QY 66 VLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 125

DB 136 TAMPAACAVEMHTVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 195

QY 126 LANDPGITVDAPAR-LKMITALTTRASSQGVGGQALDGLSVGRKLT---LPELENNMHIH 191

2Y 125 VLANDPGITVDAPARLKMITALTTRASSQGVGGQALDGLSVGRKLTPLPELENNMHIHTG 184

2b 123 CLADAPS---DTSLEIAWLOTLTAAGVAGCGGQAIDIDATGSGQTLKALEMHAHAKSG 179

2Y 185 ALIRASVNLAL-----SKPDLDTCAVKKLDHYAKICGLSVFQKDDILDIADTATLGKTQ 240

2b 180 TLMRAAVTMGALAGDPSAEDLHS-----LDTFAITGLAQVDRDDILDISSSEQLOKTA 234

2Y 241 GKDINDKPTYPALLGMAGAKQAEHLHQAVESLITGFGSEADLRLSLYIERTH 297

2b 235 GKDAIQKSTPALLSVEGAKCYLQELAERLYTQHPYGERAAPLTALARLAVERAH 291

RESULT 11

F81217

geranyltransferase NMB0261 [imported] - Neisseria meningitidis (strain MC58 serogroup B)

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C:Accession: F81217

C:Tettelein, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; H. Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzia, M.; et al. 1999-1845, 2000

A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.

A:Reference number: AB1090; MUID:20175755; PMID:10710307

C:Accession: F81217

C:Status: Preliminary

C:Molecule type: DNA

C:Residues: 1-259 <TET>

C:Cross-references: GB:AB002098; GB:AB002383; NID:G7225484; PIDN:AAF40715.1; PID:G722548

C:Experimental source: serogroup B, strain MC58

C:Genetics: NMB0261

C:Superfamily: dimethylallyltransferase

Query Match 44.2%; Score 668; DB 2; Length 259;

Best Local Similarity 52.5%; Pred. No. 1.7e-45;

Matches 137; Conservative 39; Mismatches 83; Indels 2; Gaps 1;

Y 37 MRYSLVNGKTRPLLTATGALGLPENVLDPACAVEFIHVYSLIHDDLPAMDNDL 96

b 1 MRYAALDGGKRLPMLVLAASELGEAVHEAQVQAAEMIHVYSLIHDDLPAMDNDL 60

Y 97 RGKPTCHKAYDEATAILAGDALQALAFVLANDPGITVDAPARKMITALTTRASSQGV 156

b 61 RGKPTCHKYGEATALTGALQTPDVLRSR--TELPAARQLAMSLVLRAGSGRGA 118

Y 157 GGQAIDGLSVGRKLTPLPELENNMHIHTGALIRASVNLALSKPDLDTCAVKKLDHYAKI 216

b 119 GGQAIDLANVGKQVADLQEMSLKTGALIRAAVLTGATACPDLSDBLSVLDIYAANKL 178

Y 217 GLSQVQKDDILDIADTATLGKTQKGDINDKPTYPALLGMAGAKQAEHLHQAVESLT 276

b 179 GLAQVDDVLDCEADTATLGKTAGKQANDKPTTYKMLGLEARSYAKLVLAERVALE 238

Y 277 GFQSEADLRLSLYIERTH 297

b 239 PFQDKALRLQLAEFAVARKY 259

RESULT 12

74538

geranylgeranyl pyrophosphate synthase - *Synechocystis* sp. (strain PCC 6803)

C:Alternate names: hypothetical protein slr0739

C:Species: *Synechocystis* sp.

C:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: S74538

C:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; K. Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

A:Res. 3, 109-136, 1996

Db 196 LASATSSDVSPVRVAVRAGLAKAIGTEGVAGQVVDISEGDLNDVNGLEHLEFHLH 255  
QY 182 KTGALIRASVNLAAKSPDLDTVCVAKLDHYKACIGLSFQVKDDILOIEADTATLGTQ 241  
Db 256 KTAALLLEASAVLGATVGGSDDEI--ERLRKPARCIGLLFQVVDILOVTKSSKELGRTAG 314  
QY 242 KDIDNDKPTYPALLGMAGAKAKOAEHQAQVESLTGFGSE--ADLLRELSLYIIER 295  
Db 315 KDLIADKLTYPKIMGLEKSRFAEKLNRKARDQLLGFDSKVAPLL-ALANYIAYR 369

RESULT 14  
S53722  
farnesyltransferase (EC 2.5.1.29) precursor - pepper  
N:Alternate names: geranylgeranyl diphosphate synthase; geranylgeranyl pyrophosphate syn  
C:Species: capsicum annuum (pepper)  
C:Date: 15-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 24-Sep-1999  
C:Accession: S53722  
R:Badillo, A.; Steppuhn, J.; Deruere, J.; Camara, B.; Kuntz, M.  
Plant Mol. Biol. 27, 425-428, 1995  
A:Title: Structure of a functional geranylgeranyl pyrophosphate synthase gene from Capsi  
A:Reference number: S53722; MUID:95195169; PMID:7888631  
A:Accession: S53722  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-369 <BAD>  
A:Cross-references: EMBL:X80267; NID:9643093; PIDN:CAAS6554.1; PID:9643094  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994  
C:Genetics:  
A:Gene: GGPPS  
A:Genome: nuclear  
C:Superfamily: geranyltransferase  
C:Keywords: carotenoid biosynthesis; chloroplast; chromoplast; transferase

Query Match 39.5%; Score 596; DB 2; Length 369;  
Best Local Similarity 47.1%; Pred. No. 1.3e-39;  
Matches 139; Conservative 45; Mismatches 101; Indels 10; Gaps 7;

QY 5 KAYLTVCQERVERALDARLPAENILPQTLHOAMRYSVLNGGKRTPLLTATGQALGLPE 64  
Db 79 KIYVTEKAISVNLKDEALIVKE--PHVIEAMRYSLLAGGKVRPMLCLAACELVGGNQ 136  
QY 65 NVLDAPACAVEFIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFE 124  
Db 137 ENAMAACAACAVEMIHMTSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFE 196  
QY 125 VLAND-PGIVTDAPARL-KMITALTFRASGQGMVGGQADILGVSGR-KLTLPLENNMHIH 181  
Db 197 HIVNSTAGVT---PSRIVCAVAELAKS:GTEGLVAGQVADIKCTGNASVSLTLEPIHVH 253  
QY 182 KTGALIRASVNLAAKSPDLDTVCVAKLDHYKACIGLSFQVKDDILOIEADTATLGTQ 241  
Db 254 KTAALLLESSLVGLGATVGGTNEV-EKLRKPARCIGLLFQVVDILOVTKSSSEELGRTAG 312  
QY 242 KDIDNDKPTYPALLGMAGAKAKOAEHQAQVESLTGFGS--RADLLRELSLYIIER 295  
Db 313 KOLVNDKLTYPKIMGLEKSRFAEKLNRKARDQLLGFDSKVAPLL-ALANYIAYR 367

RESULT 15  
T10452  
farnesyltransferase (EC 2.5.1.29) precursor, chloroplast - white mustard  
N:Alternate names: geranylgeranyl-diphosphate synthase; geranylgeranyl-pyrophosphate syn  
C:Species: Sinapis alba (white mustard)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 21-Jul-2000  
C:Accession: T10452  
R:Bonk, M.; Hoffmann, B.; von Lintig, J.; Schledz, M.; Al-Babili, S.; Hobeika, E.; Klein  
Eur. J. Biochem. 247, 942-950, 1997  
A:Title: Chloroplast import of four carotenoid biosynthetic enzymes in vitro reveals dif  
A:Reference number: Z17023; MUID:97433278; PMID:9286918  
A:Accession: T10452  
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA  
A:Residues: 1-366 <BON>  
A:Cross-references: EMBL:X98795; NID:91419757; PIDN:CA67330.1; PID:91419758  
C:Genetics:  
A:Gene: GGPS  
A:Genome: nuclear  
C:Superfamily: geranyltransferase  
C:Keywords: carotenoid biosynthesis; chloroplast; transferase

Query Match 39.1%; Score 590.5; DB 2; Length 366;  
Best Local Similarity 45.6%; Pred. No. 3.6e-39;  
Matches 135; Conservative 52; Mismatches 98; Indels 11; Gaps 7;

QY 6 AYLTVCQERVERALDARLPAENILPQTLHOAMRYSVLNGGKRTPLLTATGQALGLPEN 65  
Db 74 SYIIRKADSVNKALDSAVPLRE--PLKIHAMRYSLLAGGKVRPVLICIAACELVGGSE 131  
QY 66 VLDAPACAVEFIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFE 125  
Db 132 LAMPARCAVEMIHMTSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFE 191  
QY 126 LANDPGITVDAPAR-LKMITALTFRASGQGMVGGQADILGVSGRKLT---LPELENNMHIH 181  
Db 192 LASATSSSEV-SPARVVRVAVGELAKAIGTEGLVAGQVVDISEGLDNNVWGLEHLKFIHLH 250  
QY 182 KTGALIRASVNLAAKSPDLDTVCVAKLDHYKACIGLSFQVKDDILOIEADTATLGTQ 241  
Db 251 KTAALLLEASAVLGATVGGSDDEI--ERLRKPARCIGLLFQVVDILOVTKSSQBLGRTAG 309  
QY 242 KDIDNDKPTYPALLGMAGAKAKOAEHQAQVESLTGFGSE--ADLLRELSLYIIER 295  
Db 310 KDLIADKLTYPKIMGLEKSRFAEKLNRKARDQLLGFDSKVAPLL-ALANYIAYR 364

Search completed: February 29, 2004, 14:52:36  
Job time : 11.2669 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 29, 2004, 14:27:18 ; Search time 5.00562 Seconds  
(without alignments)  
3089.496 Million cell updates/sec

title: US-09-941-947A-20

effect score: 1510

sequence: 1 MSKILKYLTVCCQVERALD.....PGSEADLLRELSLYIERTH 297

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 141681 seqs, 52070155 residues

total number of hits satisfying chosen parameters: 141681

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	691.5	45.8	299	1 ISPA_ECOLI	P22939 escherichia
2	677	44.8	295	1 ISPA_HAEN	P45204 haemophilus
3	604	40.0	371	1 GGGP_ANATH	P34802 arabidopsis
4	596	39.5	369	1 GGGP_CAPAN	P80042 capsicum an
5	590.5	39.1	366	1 GGGP_SINAL	Q43133 sinapis alb
6	583	38.6	332	1 ISPA_RHISN	P55339 rhizobium s
7	578	38.3	332	1 ISPA_BRAJA	Q45220 bradyrhizob
8	577	38.2	297	1 ISPA_BACST	Q08291 bacillus st
9	575	38.1	357	1 GGGP_CATRO	Q42698 catharanthu
10	566	37.5	291	1 ISPA_MICLU	O66126 micrococcus
11	531	35.2	272	1 ISPA_BACSU	P54383 bacillus su
12	522.5	34.6	300	1 CRTE_CVAPA	P48368 cyanophora
13	501.5	33.2	294	1 ISPA_BUCAP	Q8K9A0 buchnera ap
14	501	33.2	282	1 ISPA_BUCAL	P57537 buchnera ap
15	389	26.4	289	1 CRTE_RHOCA	P17060 rhodobacter
16	369	24.4	262	1 ISPA_AQUAE	O66952 aquifex aeo
17	367	24.3	288	1 CRTE_RHOSH	P54976 rhodobacter
18	367	24.3	325	1 IDSA_METUA	O26156 methanobact
19	353.5	23.4	327	1 IDSA_METUA	Q58270 methanococc
20	353	23.4	324	1 IDSA_METUM	O53479 methanobact
21	336	22.3	302	1 CRTE_PANAN	P21684 parthea ana
22	317.5	21.0	307	1 CRTE_ERWHE	P22873 erwinia her
23	290.5	19.2	323	1 ISPB_ECOLI	P19641 escherichia
24	290.5	19.2	332	1 GGGP_SULAC	P95999 sulfolobus
25	282	18.7	330	1 GGGP_SULAC	P39464 sulfolobus
26	281	18.6	329	1 ISPB_HAEN	P44916 haemophilus
27	277.5	18.4	323	1 PREA_SVNY3	P72580 synechocyst
28	270	17.9	359	1 GGGP_XYCTU	Q50727 m probable
29	265	17.5	323	1 PREA_PORPU	P51268 porphyra pu
30	259	17.2	323	1 PREA_CTAPA	P31171 cyanophora
31	244	16.2	320	1 HEP2_BACST	P55785 bacillus st
32	226	15.0	348	1 HEP2_BACSU	P31114 bacillus su
33	215.5	14.3	378	1 DPS_SCHPO	Q43091 schizosacch

ALIGNMENTS

RESULT 1

ID	ISPA_ECOLI	STANDARD;	PRT;	299 AA.
AC	P22939;			
DT	01-AUG-1991 (Rel. 19, Created)			
DT	01-AUG-1991 (Rel. 19, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Geranyltransferase [EC 2.5.1.10] (Farnesyl-diphosphate synthase)			
DE	(PPP synthase)			
GN	ISPA OR B0421.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
[1]				
RN	SEQUENCE FROM N.A.			
RC	STRAIN=KL2;			
RC	MEDLINE=S1210228; PubMed=2089044;			
RA	Fujisaki S., Hara H., Nishimura Y., Horiuchi K., Nishino T.;			
RT	"Cloning and nucleotide sequence of the ispa gene responsible for			
RT	farnesyl diphosphate synthase activity in Escherichia coli.";			
RL	J. Biochem. 108:995-1000(1990).			
[2]				
RN	SEQUENCE FROM N.A.			
RC	STRAIN=KL2;			
RC	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shaq Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.";			
RL	Science 277:1245-1247(1997).			
[3]				
RN	SEQUENCE FROM N.A.			
RA	Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,			
RA	Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,			
RA	Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;			
RL	Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.			
CC	- - CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate			
CC	= diphosphate + trans,trans-farnesyl diphosphate.			
CC	- - SUBCELLULAR LOCATION: Cytoplasmic.			
CC	- - SIMILARITY: Belongs to the FPP/GGPP synthetase family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL; D00694; BAA00599.1; -			
DR	EMBL; AS000146; AAC73524.1; -			
DR	EMBL; U82664; AAB40177.1; -			
DR	PIR; JQ0665; JQ0665.			
DR	SWISS-2DPAGE; P22939; COLI.			

Q9t1s1 cyanidium c  
P18900 saccharomyc  
Q9wt00 m geranylge  
P56966 b geranylge  
Q95749 h geranylge  
Q82236 g geranylge  
P24322 n geranylge  
Q43315 arabidopsis  
O24242 parthenium  
Q09152 arabidopsis  
O14230 schizosacch  
P49353 zea mays (m



Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I., Vos P., Honeisel J., Zimmermann W., Wedler H., Ridley P., Langham S.-A., McCullagh B., Bilham L., Robben J., Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbosche F., Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E., Weizenegger T., Sothe G., Ramsperger U., Hilbert H., Braun M., Holzner E., Brandt A., Peters S., van Staveren M., Dirks W., Mooljman P., Klein Iankhorst R., Rose M., Hauf J., Koetter P., Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H., De Keyser A., Buysschaert C., Gielen J., Villarroel R., De Clercq R., Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S., Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R., Petrett A., Bajandream M.A., Lyne M., Benes V., Rechmann S., Borkova D., Blocker H., Scharfe M., Grimm M., Loehner T.-H., Dose S., de Haan M., Maere A., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A., Neumann S., Argiriou A., Vitale D., Liguori R., Pruvandi E., Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R., Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S., Chedford F., Cooke R., Berger C., Monfort A., Casacuberta E., Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A., Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T., Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielle C., Frishman D., Haase D., Lencke K., Mewes H.-W., Stocker S., Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K., Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L., A. Senkon M., Murray J., Sheet P., Cordes M., Abu-Threiden J., Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J., Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D., A. Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W., A. Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D., Du H., Ali J., Berghoff A., Jones K., Drone K., Colton M., Joshi C., Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C., Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R., Swaby I.K., O'Shaughnessy R., Rodriguez M., Hoffman J., Till S., Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A., Chen E., Marra M.A., Martienssen R., McCombie W.R.; "Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.";

Nature 402:769-777(1999).

FUNCTION: Catalyzes the trans-addition of the three molecules of IPP onto DMAPP to form geranylgeranyl pyrophosphate.

CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl diphosphate = diphosphate + geranyl diphosphate.

CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans-trans-farnesyl diphosphate.

CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + geranylgeranyl diphosphate.

PATHWAY: First committed step in carotenoid biosynthesis. Key enzyme in plant terpenoid biosynthesis.

SUBUNIT: Monomer.

SUBCELLULAR LOCATION: Chloroplast.

SIMILARITY: Belongs to the FPP/GGPP synthetase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L25813; AAA32797.1; -

EMBL; Z99708; CAB15803.1; -

EMBL; AL161590; CAB80347.1; -

PIR; F85434; F85434.

InterPro; IPR000092; Polyphenyl synth.

InterPro; IPR008949; Terpenoid synth.

Pfam; PF003043; Polyphenyl synth; 1.

PROSITE; PS00444; POLYPHENYL SYNTHET 2; 1.

PROSITE; PS00723; POLYPHENYL SYNTHET 1; 1.

Transferrase; isoprene biosynthesis; Carotenoid biosynthesis.

Query Match	40.0%; Score 604; DB 1; Length 371;
Best Local Similarity	45.9%; Pred. No. 7.9e-40;
Matches 136; Conservative 51; Mismatches 99; Indels 10; Gaps 16	
QY 6	AYLTVCQERVALDARLPANILPQTHQAMRVSLNGKRTPLTIVATQALGIPEN 65
DB 78	SIITKAEVLNKAELDSAYELRE--PLKTHEAMRVSLIAGGVRPVLCAACELVGGES 135
QY 66	VLDAPACAVEFTHTVYSLHDDLPAMDNDLLRRGKPTCHKAYDEATALLAGDALQALAFV 125
DB 136	TAMPACAVEMHTMSLHDDLPAMDNDLLRRGKPTNKHVFGSDVAVLAGDALLSPAFEH 195
QY 126	LANDPGITVDAPAR-LKMITALTASSQCGWVGQAIDLSGVCKULT--LPELENNMTH 181
DB 196	LSATSSDVSVPVRVAVGELAKAIGTEGLVAGQVVDISSEGLDNDVGLHLEHLEHILH 255
QY 182	KTCALIRASVNLAAISKDLDTCTVAKKLDHVAKICIGLSFOVKDDIILDIEADTATLGKTCG 241
DB 256	KTALEASAVLGAIVGGSGODEI-ERLRKFARCIQLLFOVVDIILDVTKSKKELGKTAG 314
QY 242	KDINDKPTYPALLGWAGAKAKQAEHQEVESITGFGSE--ADLLREILSLVITIER 295
DB 315	KDLIADKLTYPKIMSLKRSRFAEKLNRARDQLLGFSDKVPAPLI-ALANYIAYR 369

RESULT 4

ID	GGPP	CAPAN	STANDARD;	PRT; 369 AA.
AC	P80042;			
DT	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Geranylgeranyl pyrophosphate synthetase, chloroplast precursor (GGPP synthetase) (GGPS) [includes: Dimethylallyltransferase (EC 2.5.1.1); Geranyltransferase (EC 2.5.1.10); Farnesyltransferase (EC 2.5.1.29)]			
GN	GGPS1.			
OS	Capsicum annuum (Bell pepper).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;			
OC	lamids; Solanales; Solanaceae; Capsicum.			
OX	NCBI_TaxID=4072;			
RI	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Lamuyo; TISSUE=Fruit;			
RC	MEDLINE=93272043; PubMed=1303794;			
RA	Kuntz M., Roemer S., Suire C., Hugueney P., Weil J.H., Schantz R.,			
RA	Camara B.;			
RT	"Identification of a cDNA for the plastid-located geranylgeranyl			
RT	pyrophosphate synthase from Capsicum annuum: correlative increase in			
RT	enzyme activity and transcript level during fruit ripening.";			
RL	Plant J. 2:25-34 (1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=cv. Yolo Wonder;			
RC	MEDLINE=95195169; PubMed=7888631;			
RA	Badillo A., Steppuhn J., Deruere J., Camara B., Kuntz M.;			
RT	"Structure of a functional geranylgeranyl pyrophosphate synthase gene			
RT	from Capsicum annuum.";			
RL	Plant Mol. Biol. 27:425-428 (1995).			
CC	-!- FUNCTION: Catalyzes the trans-addition of the three molecules of			
CC	IPP onto DMAPP to form geranylgeranyl pyrophosphate.			
CC	-!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl			
CC	diphosphate = diphosphate + geranyl diphosphate			
CC	-!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate			



01-NOV-1997 (Rel. 35, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
Probable geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase) (FPP synthase).  
Rhizobium sp. (strain NGR234).  
Plasmid sym pNGR234a.  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
NCBI\_TaxID=394;  
SEQUENCE FROM N.A.  
MEDLINE=97305956; PubMed=9163424;  
Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A., Perret X.;  
Nature 387:394-401(1997).  
"Molecular basis of symbiosis between Rhizobium and legumes";  
NATURE 387:394-401(1997).  
-!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans,trans-farnesyl diphosphate.  
-!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
-!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; AE000082; AAB91752.1; .  
PIR; T10875; T10875.  
InterPro; IPR000092; Polyrenyl synt.  
InterPro; IPR008949; Terpenoid synth.  
Pfam; PF00348; polyrenyl synt; 1.  
PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.  
PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.  
Transferase; Isoprene biosynthesis; Plasmid.  
SEQUENCE 332 AA; 34688 MW; EE68C1547449AA87 CRC64;  
Query Match 38.6%; Score 583; DB 1; Length 332;  
Best Local Similarity 45.7%; Pred. No. 3e-38;  
Matches 133; Conservative 46; Mismatches 98; Indels 14; Gaps 4;  
13 ERVERALDARLPARNILPQTLHQAAMRYSLVNGGKTRPLTYATGQALGPENVLDPAC 72  
42 KRVEALRLCAEDHGETELMAAMRYATLFGKTRALLCLAGALADTPAHILDDVGA 101  
73 AVEFIHVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLA 132  
102 AIENMHACTLVHDDLPAMDNDLRRGLPTVHVKFGTEATILVGDALQAHAFLLTA 156  
133 TVDAPA--RLKMITALFRASGQMGVGGQALDLSGVKRLTLELENMHIHKTGALIRAS 190  
157 SLDPAGCNRLAVRELAAQVSAEGAGGQAMDSLGVKVELDRIIVAMRMKCALVRAS 216  
191 VN---LAALSKPDLDTCAKKLHYAKICGLSPQVKKDDILDIEDATATLTKGTQGDIND 247  
217 VRMGALCAIAEADAATLYCALDRYSACFGLAQVDDILDADATATLTKGTGKDAQA 276  
248 KPTYPALLGWAGAKQAKELHEQAVESLITGFSGEAD-----LLRELSLYIE 294  
277 KPTCASIMGLQAQKAFALDILLCSEAGEAIPGPRAEALQMQRASAYLFK 327  
RESULT 7  
\_SPA\_BRAJA  
\_D\_ISPA\_BRAJA STANDARD; PRT; 332 AA.  
Q45220;  
01-NOV-1997 (Rel. 35, Created)  
10-NOV-1997 (Rel. 35, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
Probable geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate

synthase) (FPP synthase).  
FPPS OR BUR2148.  
Bradyrhizobium japonicum.  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium.  
NCBI\_TaxID=375;  
SEQUENCE FROM N.A.  
STRAIN=USDA 110;  
Tully R.E., Keister D.L.;  
"Cloning and mutagenesis of a cytochrome P-450 locus from Bradyrhizobium japonicum that is expressed anaerobically and symbiotically";  
Appl. Environ. Microbiol. 59:4136-4142(1993).  
SEQUENCE FROM N.A.  
STRAIN=USDA 110;  
MEDLINE=98322110; PubMed=9655913;  
Tully R.E., van Berkum P., Lovins K.W., Keister D.L.;  
"Identification and sequencing of a cytochrome P450 gene cluster from Bradyrhizobium japonicum";  
Bradyrhizobium japonicum";  
Biochim. Biophys. Acta 1398:243-255(1998).  
SEQUENCE FROM N.A.  
STRAIN=USDA 110;  
MEDLINE=22484998; PubMed=12597275;  
Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T., Sasamoto S., Watanabe A., Idegawa K., Iriyuchi M., Kawashina K., Kohara M., Matsumoto M., Shimpou S., Tsuruoka H., Wada T., Yamada M., Tabata S.;  
"Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110";  
DNA Res. 9:189-197(2002).  
-!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate = diphosphate + trans,trans-farnesyl diphosphate.  
-!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
-!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; U12678; AAC28894.1; .  
EMBL; AP005942; SACC47413.1; ALT\_INIT.  
PIR; I40213; I40213.  
InterPro; IPR000092; Polyrenyl synt.  
InterPro; IPR008949; Terpenoid synth.  
Pfam; PF00348; polyrenyl synt; 1.  
PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.  
PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.  
Transferase; Isoprene biosynthesis; Complete proteome.  
SEQUENCE 332 AA; 34638 MW; 47644FAF0220CE4C CRC64;  
Query Match 38.3%; Score 578; DB 1; Length 332;  
Best Local Similarity 47.3%; Pred. No. 7.4e-38;  
Matches 133; Conservative 40; Mismatches 98; Indels 10; Gaps 4;  
13 ERVERALDARLPARNILPQTLHQAAMRYSLVNGGKTRPLTYATGQALGPENVLDPAC 72  
42 KRVEALRLCAEDHGETELMAAMRYATLFGKTRALLCLAGALADTPAHILDDVGA 101  
73 AVEFIHVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLA--NDP 130  
102 AIENMHACTLVHDDLPAMDNDLRRGLPTVHVKFGTEATILVGDALQAHAFLLTA 161  
131 GITVDAPARLKMITALFRASGQMGVGGQALDLSGVKRLTLELENMHIHKTGALIRAS 190  
162 G---DSP--IALVRELAQAVSAEGAGGQALDLSLVGKVELDRIIVAMRMKCALVRAS 216

QY 191 VMLAISKEDLTCVAK---KLDHYAKICIGISFQVKDDILDEADTATLGTQKQDIND 247  
DB 217 VRMGALCAVGNVNAARALYCALDHYSAFGLAQVDDILSVTADTAALGKTCKGDAQA 276  
QY 248 KTYTPALLGMAGAKOAKOELHEQAVESITGFGSEADLLREL 288  
DB 277 KTCASIMELQARQFALDLDLADAGEATAPGPRALQAL 317

RESULT 8  
ISPA BACST  
ID ISPA BACST STANDARD; PRT; 297 AA.  
AC Q0831; Q53435; Q53436; Q53437; Q53438;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)  
DE (fpp synthase).  
OS Bacillus stearothermophilus.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.  
OX NCBI\_TaxID=1422;  
RN [1]  
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RP STRAIN=ATCC 10149;  
RX MEDLINE=93252758; PubMed=8486607;  
RA Koyama T., Obata S., Osabe M., Takeshita A., Yokoyama K.,  
RA Uchida M., Nishino T., Ogura K.;  
RT "Structural and functional roles of the cysteine residues of Bacillus  
RT stearothermophilus farnesyl diphosphate synthase of Bacillus  
RT stearothermophilus; molecular cloning, sequence determination,  
RT overproduction, and purification.";  
RN J. Biochem. 113:355-363(1993).  
RN [2]  
MUTAGENESIS OF Cysteine Residues.  
RP MEDLINE=95001990; PubMed=7918490;  
RX Koyama T., Obata S., Saito K., Takeshita-Koike A., Ogura K.;  
RA "Structural and functional roles of the cysteine residues of Bacillus  
RT stearothermophilus farnesyl diphosphate synthase.";  
RL Biochemistry 33:12644-12648(1994).  
CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate  
CC = diphosphate + trans,trans-farnesyl diphosphate.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D13293; BAA02551.1; -  
CC DR EMBL; S72629; AAB32272.1; -  
CC DR EMBL; S72630; AAB32273.2; ALT\_SEQ.  
CC DR EMBL; S72633; AAB32274.1; -  
CC DR EMBL; S72635; AAB32275.2; ALT\_SEQ.  
CC PIR; JX0257; JX0257.  
CC InterPro; IPR000092; Polyrenyl\_synth.  
CC InterPro; IPR008949; Terpenoid\_synth.  
CC Pfam; PF00348; polyrenyl\_synth; 1.  
CC PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.  
CC PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.  
KW Transferase; Isoprene biosynthesis.  
FT MUTAGEN 289 289 C->F; S: NO LOSS OF ACTIVITY.  
FT MUTAGEN 289 289 C->P; S: NO LOSS OF ACTIVITY.  
SQ SEQUENCE 297 AA; 32310 MW; 0F921C3F029EEB6 CRC64;  
Query Match 38.2%; Score 577; DB 1; Length 297;  
Best Local Similarity 44.6%; Pred. No. 7.7e-38;  
Matches 133; Conservative 42; Mismatches 114; Indels 8; Gaps 3;  
QY 4 LKAYLTCGVERALD---ABLPALNTLPQTLHQAEMYSVLNGGKTRPLLTATGQAL 60

DB 6 VEQFLNEQQAQVETALSRYIERLEG----PAKIKCAMAYSLEAGGKIRPILLSTVRAL 61  
QY 61 GLPENVLDPACAVAFIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDAFALAGDALQA 120  
DB 62 GDPAVELPVCACAEIEMINTYSLIHDDLPAMDNDLRRGKPTNHKVFGEAMAILAGDGLUT 121  
QY 121 LAFEVLANDPGITVDAPARLKMITALTTRASGQGVGGCAIDLAGSVGRKLTLPLENNHI 180  
DB 122 YAFQLITEIDDERIPPSVRLRIERLAKAAGPGNVAGQAADMEGKTLTLELYIHR 181  
QY 181 HMTGALIRASVNLAAKSPDLDTCVAKKLDHYAKICIGISFQVKDDILDEADTATLGTQ 240  
DB 182 HMTGKMLQVSVHAGALI--GGADARQTRDELDEFAHLGLAFQIRDDILDIEGAEEKIGEPV 240  
QY 241 GKDINDKTTYBALLGMAGAKOAKOELHEQAVESITGFGSEADLLRELSLYIERTH 297  
DB 241 GSDQNNKATYPALLSLAGAKELAFHIAAQRHLRNADVGAALAYICELVAARDH 297

RESULT 9  
GGPP CATRO  
ID GGPP CATRO STANDARD; PRT; 357 AA.  
AC Q42698;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DE Geranylgeranyl pyrophosphate synthetase, chloroplast precursor (GGPP  
DE synthetase) (GGPS) [includes: Dimethylallyltransferase (EC 2.5.1.1);  
DE Geranyltransferase (EC 2.5.1.10); Farnesyltransferase  
DE (EC 2.5.1.29)].  
GN GGPS1 OR GGC1.  
OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).  
OC Sukkaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
OC Lamiales; Gentianales; Apocynaceae; Rauvolfioideae; Vinceae;  
OC Catharanthus.  
OX NCBI\_TaxID=4058;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=cv. G. Don c20;  
RA Bantignies B., Liboz T., Ambid C.;  
RT "Nucleotide sequence of a Catharanthus roseus geranylgeranyl  
RT pyrophosphate synthase gene.";  
RL (in) Plant Gene Register PGR95-119.  
CC -!- FUNCTION: Catalyzes the trans-addition of the three molecules of  
CC IPP onto DMAPP to form geranylgeranyl pyrophosphate.  
CC -!- CATALYTIC ACTIVITY: Dimethylallyl diphosphate + isopentenyl  
CC diphosphate = diphosphate + geranyl diphosphate.  
CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate  
CC = diphosphate + trans,trans-farnesyl diphosphate.  
CC -!- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl  
CC diphosphate = diphosphate + geranylgeranyl diphosphate.  
CC -!- PATHWAY: First committed step in carotenoid biosynthesis. Key  
CC enzyme in plant terpenoid biosynthesis.  
CC -!- SUBCELLULAR LOCATION: Chloroplast.  
CC -!- SIMILARITY: Belongs to the FPP/GGPP synthetase family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X32893; CAA63486.1; -  
CC PIR; T09966; T09966.  
CC InterPro; IPR000092; Polyrenyl\_synth.  
CC InterPro; IPR008949; Terpenoid\_synth.  
CC Pfam; PF00348; polyrenyl\_synth; 1.  
CC PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.  
CC PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.  
KW Transferase; Isoprene biosynthesis; Carotenoid biosynthesis;



W Chloroplast; Transit peptide. CHLOROPLAST (POTENTIAL).  
T TRANSIT 1 40  
T CHAIN 41 357 GERANYLGERANYL PYROPHOSPHATE SYNTHETASE.  
Q SEQUENCE 357 AA; 38786 MW; 81C52FDDAL506FAS CRC64;  
Query Match 38.1%; Score 575; DB 1; Length 357;  
Best Local Similarity 47.1%; Pred. No. 1.4e-37;  
Matches 139; Conservative 43; Mismatches 103; Indels 10; Gaps 7;  
Y 5 KAVITVCOEERVAL-DARLPAENILPQTHQAMRYSLVNGKRTPLTYATGOALGUP 63  
b 67 KAYIGKANSVNNKALDEAVLVRE---PLKHESMYSLLAGGRVPMUCIAACELFGGT 123  
Y 64 ENVLDAPACAVEFIHVYSLIHDDLPAWDDLRGKPTCHKAYDEATALLAGDALQALAF 123  
b 124 ESWAPSACAVEMHTWSLWHDLDLPKWDNDLRGKPTNKHVFGEDVAVLAGDALLAPAF 183  
Y 124 EVLAN-DPGITVDAPARKMITALTTRASSOGWVGQAIDLSVG-RKLTPELENMHIH 181  
b 184 EHTATATKGVSSERIVR--VVGELAKICIGSEGVAGQVVDVCSBGIAADVGLHLEFIH 241  
Y 182 KTGALRASVNLAAKSPDLDTCAVKLDHYAKICIGLSFQVKKDILDIEADTATLTKTOG 241  
b 242 KTAALLEGSVLGAIVGANDDEIS-KLKFPARCIGLLFQVDDILDVTKSQELGKTAG 300  
Y 242 KQINDKPTYPALLGMAGAKQAKQELHQAESLTGFGSE-ADLRLRLSLYIER 295  
b 301 KDLVADKVTYKLLGIDKSRFAKLNREAEQLEAFDPEKAAPLIALANYIAYR 355  
RESULT 10  
SPA\_MICLU  
D ISPA\_MICLU STANDARD; PRT; 291 AA.  
C 066126;  
T 15-JUL-1999 (Rel. 38, Created)  
T 15-JUL-1999 (Rel. 38, Last sequence update)  
T 10-OCT-2003 (Rel. 42, Last annotation update)  
E Geranyltransterase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)  
F (FPP synthase).  
N FPS.  
S Micrococcus luteus (Micrococcus lysodeikticus).  
C Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
C Micrococcineae; Micrococcaceae; Micrococcus.  
X NCBI\_TaxID=1270;  
N [1]  
P SEQUENCE FROM N.A.  
C STRAIN=B-P 26;  
X MEDLINE=98175686; PubMed=9515931;  
A Shimizu N., Koyama T., Ogura K.;  
T "Molecular cloning, expression, and characterization of the genes  
encoding the two essential protein components of Micrococcus luteus  
B-P 26 hexaprenyl diphosphate synthase.";  
L J. Bacteriol. 180:1578-1581 (1998).  
C -/- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate  
= diphosphate + trans,trans-farnesyl diphosphate.  
C -/- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
C -/- SIMILARITY: Belongs to the FPP/G3PP synthetase family.  
C This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
C EMBL; AB003187; BAA25265.1; -  
R InterPro; IPR000092; Polyprenyl synth.  
R InterPro; IPR008949; Terpenoid synth.  
R Pfam; PF00348; polyprenyl synth. 1.  
R PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.  
R PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.  
W Transferase; Isoprene biosynthesis.

SQ SEQUENCE 291 AA; 32362 MW; 797125AB71B5674A CRC64;  
Query Match 37.5%; Score 566; DB 1; Length 291;  
Best Local Similarity 45.5%; Pred. No. 5.4e-37;  
Matches 130; Conservative 44; Mismatches 92; Indels 20; Gaps 5;  
QY 16 ERAIDARLPAENILPQTHQAMRYSLVNGKRTPLTYATGOAL-----GLPENVLA 69  
b 18 ESLNKHVPAQS---RLHEAINVSLSAGGRRLPLVLTLDLSLGNHAGLPPGI--- 70  
QY 70 PACAVEFIHVYSLIHDDLPAWDDLRGKPTCHKAYDEATALLAGDALQALAFVLAN 129  
b 71 ---ALEMIHTYSLIHDDLPAWDDLRGKPTCHKAYDEATALLAGDALQALAFVLAN 126  
QY 130 PGITVDAPARKMITALTTRASSOGWVGQAIDLSVG-RKLTPELENMHIHKTGALIRA 189  
b 127 --TQLNREIKLSLNLSTASGSGVVGQMDMQEHKTLINELERHIHKTGALIRA 184  
QY 190 SVNLAAKSPDLDTCAVKLDHYAKICIGLSFQVKKDILDIEADTATLTKTOGKIDNDKP 249  
b 185 AIVSAGIIM-NFNDAQTEQLNIIGKNVGLMFQIKDDILDVEGSEFENIKTVGSDLNDRKS 243  
QY 250 TYPALLGMAGAKQAKQELHQAESLTGFGSEADLRLRLSLYIER 295  
b 244 TYVSLGLSEASKOLLNDKLTETTYDAKLTQINDNKLTLITYIVER 289  
RESULT 11  
ISPA\_BACSU  
D ISPA\_BACSU STANDARD; PRT; 272 AA.  
AC P54383;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Geranyltransterase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)  
DS (FPP synthase).  
GN YQID OR BSU24289.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
N [1]  
P SEQUENCE FROM N.A.  
C STRAIN=168 / JH642;  
X MEDLINE=97124195; PubMed=8969508;  
R Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,  
RA Kobayashi Y.;  
T "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of  
the Bacillus subtilis genome containing the skin element and many  
RT sporulation genes.";  
RL Microbiology 142:3103-3111 (1996).  
R [2]  
P SEQUENCE FROM N.A.  
C STRAIN=168;  
X MEDLINE=98044033; PubMed=9384377;  
R Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borries R., Boursier L., Brans A., Braun M., Brignall S.C., Bron S.,  
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.F.,  
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Gim S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,  
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,



RESULT 13  
 ID ISPA\_BUCAP STANDARD; PRT; 294 AA.  
 AC Q8K9A0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)  
 CE (FPP synthase)  
 EN ISPA OR BUSG449.  
 OS Buchnera aphidicola (subsp. Schizaphis graminum).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Buchnera.  
 CX NCBI\_TaxID=98794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22084549; PubMed=12089438;  
 RA Tamás I., Klasson L., Canbeek B., Naeslund A. K., Eriksson A.-S.,  
 RT Wernegreen J. J., Sandstrom J. P., Moran N. A., Anderson S. C. R.;  
 RL "50 million years of genomic stasis in endosymbiotic bacteria."; Science 296:2376-2379(2002).  
 CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate  
 = diphosphate + trans,trans-farnesyl diphosphate.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the FPP/GPPP synthetase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AB014121; AAMG7992.1; -  
 DR InterPro; IPR000092; Polyprenyl synt.  
 DR InterPro; IPR008949; Terpenoid synth.  
 DR Pfam; PF00348; Polyprenyl synt; 1.  
 DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.  
 DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.  
 KW Transferase; Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 294 AA; 33880 MW; 7D7D901C7213ED0A CRC64;  
 Query Match 33.2%; Score 501.5; DB 1; Length 294;  
 Best Local Similarity 42.2%; Pred. No. 5-32;  
 Matches 117; Conservative 49; Mismatches 98; Indels 13; Gaps 5;  
 23 LPAENILPQTLHQMRYSVINGKRTPLITYATGQALGPENVLAPACAVEFIHVYSL 82  
 25 LPFQD---SVLFRAMKYSTLSGGKRLRACLIYATGETFQVNI AALDVISAAVELVHSYSL 81  
 83 IHDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLAN--PGITVDAPARL 140  
 82 IHDLPICDNDYFRGKISCHKIYGENFALLAGDALQALAFVLAN--VHDISIRL 139  
 141 KMTALTTRASSQGWGQALDGSVGRKLTLPLENMHIHTGALIRASVNLAALESKD 200  
 140 KMAEFSNAIGSYGCMGQMLDEKEREKINI SELEKINLYKTAFIRCSIRLAYFASNN 199  
 201 LDTCAVKLQHYAKICGLSVQKDDILDLEADTATLTKTGKQDINDKPTYPALLQWAGA 260  
 200 FSKVELFILDKFSVIGLAFQIQDILLKNDIKLESKNNK---TKNTYPLLIGLKKS 255  
 261 KQKAEHLHQAVESLTGF--GSEADLRLRELSYIIER 295  
 256 KIKIKELYKEAFFTEILKKNFNVLKLTQFMKR 292  
 RESULT 14  
 ID ISPA\_BUCAP STANDARD; PRT; 282 AA.  
 AC Q8K9A0;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)  
 CE (FPP synthase)  
 EN ISPA OR BUSG449.  
 OS Buchnera aphidicola (subsp. Schizaphis graminum).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Buchnera.  
 CX NCBI\_TaxID=98794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22084549; PubMed=12089438;  
 RA Tamás I., Klasson L., Canbeek B., Naeslund A. K., Eriksson A.-S.,  
 RT Wernegreen J. J., Sandstrom J. P., Moran N. A., Anderson S. C. R.;  
 RL "50 million years of genomic stasis in endosymbiotic bacteria."; Science 296:2376-2379(2002).  
 CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate  
 = diphosphate + trans,trans-farnesyl diphosphate.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the FPP/GPPP synthetase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AB014121; AAMG7992.1; -  
 DR InterPro; IPR000092; Polyprenyl synt.  
 DR InterPro; IPR008949; Terpenoid synth.  
 DR Pfam; PF00348; Polyprenyl synt; 1.  
 DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.  
 DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.  
 KW Transferase; Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 282 AA; 33880 MW; 7D7D901C7213ED0A CRC64;  
 Query Match 33.2%; Score 501; DB 1; Length 282;  
 Best Local Similarity 41.9%; Pred. No. 6-32;  
 Matches 112; Conservative 56; Mismatches 91; Indels 8; Gaps 4;  
 13 EREVERALDARL-PAENTLP---QTLHQMRYSVINGKRTPLITYATGQALGPENVL 68  
 8 ERYKRNQKLFYTLNQLPFQKXSLKAMKYSVFGSKRLRSLSLTYSTGDFVKVNTITLD 67  
 69 APACAVEFIHVYSYLHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLAN 128  
 68 VISTAIFHSYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLAN 127  
 129 PGITVDAPARLKMITALTRASSQGWGQALDGSVGRKLTLPLENMHIHTGALIR 188  
 128 SFMPNVSNLKRIMSELSYISGSGCMGQNLDEAKKDVNLSELEIINLYKTSLMR 187  
 189 ASVNLAALESKDPTLDTCAVKLQHYAKICGLSVQKDDILDLEADTATLTKTGKQDINDK 248  
 188 SAVELVVFSSNNFSSKSLTSLDLPFSITSLGAFQIQDILLDFKXDSV---KTDNKKLIK-K 243  
 249 PTYPALLQWAGAKAKQAEHLHQAVESL 275  
 244 HTYPLIIGLDESRRKKIKQLHKKSFLAL 270  
 RESULT 15  
 ID CRTE\_RHOCA STANDARD; PRT; 289 AA.  
 AC F17060;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)

P57537;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Geranyltransferase (EC 2.5.1.10) (Farnesyl-diphosphate synthase)  
 CE (FPP synthase);  
 EN ISPA OR BU465.  
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum  
 symbiotic bacterium).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Buchnera.  
 CX NCBI\_TaxID=118099;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=Tokyo 1998;  
 RX MEDLINE=20445173; PubMed=10993077;  
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "Genome sequence of the endocellular bacterial symbiont of aphids  
 Buchnera sp. APS."; Nature 407:81-86(2000).  
 CC -!- CATALYTIC ACTIVITY: Geranyl diphosphate + isopentenyl diphosphate  
 = diphosphate + trans,trans-farnesyl diphosphate.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to the FPP/GPPP synthetase family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AP001119; BAB13162.1; -  
 DR InterPro; IPR000092; Polyprenyl synt.  
 DR InterPro; IPR008949; Terpenoid synth.  
 DR Pfam; PF00348; Polyprenyl synt; 1.  
 DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.  
 DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.  
 KW Transferase; Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 282 AA; 32331 MW; 228B31A6BCD6445 CRC64;  
 Query Match 33.2%; Score 501; DB 1; Length 282;  
 Best Local Similarity 41.9%; Pred. No. 6-32;  
 Matches 112; Conservative 56; Mismatches 91; Indels 8; Gaps 4;  
 13 EREVERALDARL-PAENTLP---QTLHQMRYSVINGKRTPLITYATGQALGPENVL 68  
 8 ERYKRNQKLFYTLNQLPFQKXSLKAMKYSVFGSKRLRSLSLTYSTGDFVKVNTITLD 67  
 69 APACAVEFIHVYSYLHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLAN 128  
 68 VISTAIFHSYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLAN 127  
 129 PGITVDAPARLKMITALTRASSQGWGQALDGSVGRKLTLPLENMHIHTGALIR 188  
 128 SFMPNVSNLKRIMSELSYISGSGCMGQNLDEAKKDVNLSELEIINLYKTSLMR 187  
 189 ASVNLAALESKDPTLDTCAVKLQHYAKICGLSVQKDDILDLEADTATLTKTGKQDINDK 248  
 188 SAVELVVFSSNNFSSKSLTSLDLPFSITSLGAFQIQDILLDFKXDSV---KTDNKKLIK-K 243  
 249 PTYPALLQWAGAKAKQAEHLHQAVESL 275  
 244 HTYPLIIGLDESRRKKIKQLHKKSFLAL 270  
 CRTE\_RHOCA  
 ID CRTE\_RHOCA STANDARD; PRT; 289 AA.  
 AC F17060;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)  
Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29) (GGPP synthetase) (Farnesyltransferase).  
CITE.  
Rhodobacter capsulatus (Rhodospseudomonas capsulata).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;  
OC Rhodobacteraceae; Rhodobacter.  
NCBI\_TaxID=1061;  
[1]  
SEQUENCE FROM N.A.  
RP STRAIN=SB1003 / St. Louis, and BEC404;  
RX MEDLINE=89313663; PubMed=2747617;  
RA Armstrong G.A., Alberti M., Leach F., Hearst J.E.;  
RT "Nucleotide sequence, organization, and nature of the protein products of the carotenoid biosynthesis gene cluster of Rhodobacter capsulatus";  
RT Mol. Gen. Genet. 216:254-268 (1989).  
RL  
CC -1- CATALYTIC ACTIVITY: Trans-trans-farnesyl diphosphate + isopentenyl diphosphate = diphosphate + geranylgeranyl diphosphate.  
CC -1- PATHWAY: Carotenoid and chlorophyll biosynthesis.  
CC -1- SIMILARITY: Belongs to the FPP/GGPP synthetase family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; X52291; CAA36538.1; -;  
CC EMBL; Z11165; CAA77545.1; -;  
CC PIR; S04407; S04407.  
DR InterPro; IPR000092; Polyprenyl synth.  
DR InterPro; IPR008949; Terpenoid synth.  
DR Pfam; PF00348; polyprenyl synth; 1.  
DR PROSITE; PS00444; POLYPRENYL SYNTHET 2; 1.  
DR PROSITE; PS00723; POLYPRENYL SYNTHET 1; 1.  
KW Photosynthesis; Chlorophyll biosynthesis; Carotenoid biosynthesis; isoprene biosynthesis; transferase.  
KW SEQUENCE 289 AA; 30043 MW; CF483A26ECNA9C859 CRC64;  
Query Match 26.4%; Score 399; DB 1; Length 289;  
Best Local Similarity 37.4%; Pred. No. 5.5e-24;  
Matches 108; Conservative 46; Mismatches 107; Indels 28; Gaps 9;  
Qy 13 ERVERALDARLPENI--LPQTLHQMRYSLVNGGKTRPLITYATGQALG--LPENVLD 68  
Db 5 KRIESALVKALSPEALGESPLLAALPYGVFPFGARIRPTILVSVLACGDDCPA-VTD 63  
Qy 69 APACAVEFIHWYSLIHDDLPAMDNDLIRGKPTCKAYDEATAILAGDALQALAEVLAN 128  
Db 64 AAVALLELMHCASLVHDDLPADFNDIRGKPSLHKAYNEPLAVLAGDSLLIRGFEVLA- 122  
Qy 129 DPGITVDAPARKVITALTASGQ-GWGGQALDLSGVGRKLTLPLENMHIHTGALI 187  
Db 123 DVG-AVNPDRALKLISKLGQSGARGICAGQAWSESE-----KVDLAAYHQAKTGALF 175  
Qy 188 RASVNLAAAL-----SKPDLDTVCVAKKLHVYAKCIGLSFOVKDDILDIEADTATLGKTOGK 242  
Db 176 IATQMGAIAGYEAEWFDLGR-----IGSAFQIADDLKDALMSAEMKPGAQ 226  
Qy 243 DIDNDKFTYFALLGMAGAKQAKQELHEQAVESLTGFGSRADLLRELSLY 291  
Db 227 DIANERNVAKTMGIEGARKHLQDLVLAGATIASIPSCPGAKLAQMVOLY 275

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

run on: February 29, 2004, 14:51:24 ; Search time 19.7191 Seconds

(without alignments)

3180.293 Million cell updates/sec

Title: US-09-941-947A-20

Perfect score: 1510

Sequence: 1 MSKLKAYLVQCEVERALD.....FQSEADLLRELSUYIERTH 297

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1510	100.0	297	9	US-09-934-903-14
2	1510	100.0	297	9	US-09-934-868-72
3	1510	100.0	297	10	US-09-941-947A-20
4	831	55.0	295	9	US-09-815-242-11971
5	791.5	52.4	291	15	US-10-369-493-8506
6	782	51.8	295	15	US-10-369-493-13999
7	719.5	47.6	294	15	US-10-369-493-4670
8	719.5	47.6	294	15	US-10-369-493-7429
9	715	47.4	291	15	US-10-369-493-15449
10	707	46.8	284	15	US-10-369-493-15617
11	707	46.8	284	15	US-10-369-493-16200
12	705.5	46.7	299	9	US-09-815-242-14084
13	691.5	45.8	299	9	US-09-815-242-10069
14	691.5	45.8	299	15	US-10-369-493-732
15	690	45.7	306	15	US-10-369-493-10187

Sequence 480, App  
Sequence 11239, A  
Sequence 17527, A  
Sequence 9290, Ap  
Sequence 9246, Ap  
Sequence 21173, A  
Sequence 9513, Ap  
Sequence 10617, A  
Sequence 18, Appl  
Sequence 20893, A  
Sequence 2608, Ap  
Sequence 18752, A  
Sequence 19922, A  
Sequence 12201, A  
Sequence 18010, A  
Sequence 26, Appl  
Sequence 159, Appl  
Sequence 17358, A  
Sequence 22, Appl  
Sequence 9710, Ap  
Sequence 45, Appl  
Sequence 2, Appl  
Sequence 16549, A  
Sequence 16943, A  
Sequence 20569, A  
Sequence 23201, A  
Sequence 64, Appl  
Sequence 54, Appl  
Sequence 5239, Ap  
Sequence 12583, A

682.5 45.2 296 15 US-10-369-493-480  
677 44.7 295 9 US-09-815-242-11239  
675 44.7 291 15 US-10-369-493-17527  
672 44.5 291 15 US-10-369-493-9290  
671 44.4 256 15 US-10-369-493-9546  
659.5 44.3 296 15 US-10-369-493-21173  
661 43.8 276 15 US-10-369-493-9513  
617 40.9 292 15 US-10-369-493-10617  
610 40.4 369 13 US-10-108-915-18  
608 40.3 290 15 US-10-369-493-20893  
606.5 40.2 302 15 US-10-369-493-2608  
586 38.7 309 15 US-10-369-493-18752  
585 38.7 310 15 US-10-369-493-19922  
576 38.1 321 15 US-10-369-493-12201  
574 38.0 294 15 US-10-369-493-18010  
572 37.9 367 13 US-10-108-915-26  
571.5 37.8 287 14 US-10-166-225A-159  
565 37.4 294 15 US-10-369-493-17358  
565 37.4 350 13 US-10-108-915-22  
559 37.0 261 15 US-10-369-493-9710  
552.5 36.6 316 13 US-10-108-915-45  
551 36.5 377 9 US-09-934-778-2  
548.5 36.3 297 15 US-10-369-493-16549  
541.5 35.9 294 15 US-10-369-493-16943  
538.5 35.7 296 15 US-10-369-493-20569  
531 35.2 272 15 US-10-369-493-23201  
516.5 34.2 287 9 US-09-925-637-64  
516.5 34.2 287 14 US-10-084-205-64  
512.5 33.9 288 9 US-09-815-242-5239  
512.5 33.9 293 9 US-09-815-242-12583

#### ALIGNMENTS

#### RESULT 1

US-09-934-903-14  
; Sequence 14, Application US/09934903  
; Patent No. US20020102690A1  
; GENERAL INFORMATION:  
; APPLICANT: Koffas, Mattheos  
; APPLICANT: Odum, J. Martin  
; APPLICANT: Schenzle, Andreas J.  
; APPLICANT: No. US20020102690A1ton, Kelley C.  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Rouviere, Pierre  
; APPLICANT: Picataggio, Stephen  
; APPLICANT: Cheng, Qiong  
; TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production  
; FILE REFERENCE: CIL646 US NA  
; CURRENT APPLICATION NUMBER: US/09/934,903  
; CURRENT FILING DATE: 2001-08-22  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR FILING DATE: 60/229,907  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 14  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Methylobionas 16a  
; FEATURE:  
; OTHER INFORMATION: Amino acid sequences encoded by ORF7  
US-09-934-903-14

Query Match 100.0%; Score 1510; DB 9; Length 297;  
Best Local Similarity 100.0%; Pred. No. 4.4e-152;  
Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSKLKAYLVQCEVERALDARLPAENILPQTLHQAMRYSLVNGGKTRPLLYATVQAL 60  
DB 1 MSKLKAYLVQCEVERALDARLPAENILPQTLHQAMRYSLVNGGKTRPLLYATVQAL 60  
QY 61 GLPENLIDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQA 120

Db 61 GLPENVLDPACAVAFIHYVSLIHDLPAMDNDLRRGKPTCHKAYDEATLADALQA 120  
QY 121 LAPEVLNDPGITVDAPARLKMITALTRASQGVGGQADLGSVGRKLTLPLENWHI 180  
Db 121 LAPEVLNDPGITVDAPARLKMITALTRASQGVGGQADLGSVGRKLTLPLENWHI 180  
QY 181 HKTGALIRASVNLAAALSKPDLDTCAVAKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240  
Db 181 HKTGALIRASVNLAAALSKPDLDTCAVAKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240  
QY 241 GKDINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRELSLYIERTH 297  
Db 241 GKDINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRELSLYIERTH 297

RESULT 2

US-09-934-868-72  
; Sequence 72, Application US/09934868  
; Patent No. US20020137190A1  
; GENERAL INFORMATION:  
; APPLICANT: Koffas, Mattheos  
; APPLICANT: Odom, James M  
; APPLICANT: Schenzle, Andreas J  
; TITLE OF INVENTION: DENITRIFYING METHANOTROPHIC BACTERIAL STRAIN  
; FILE REFERENCE: CL1596 US NA  
; CURRENT APPLICATION NUMBER: US/09/934,868  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/229,858  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 81  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 72  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Methylobionas 16a  
; FEATURE:  
; OTHER INFORMATION: Amino acid sequences encoded by ORF7 - ISPA  
US-09-934-868-72

Query Match 100.0%; Score 1510; DB 9; Length 297;  
Best Local Similarity 100.0%; Pred. No. 4.4e-152;  
Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKLAAYLTVCQVERALDARLPAENILPQTLHQAMRYSVLNGGKTRPLTYATGOAL 60  
Db 1 MSKLAAYLTVCQVERALDARLPAENILPQTLHQAMRYSVLNGGKTRPLTYATGOAL 60  
QY 61 GLPENVLDPACAVAFIHYVSLIHDLPAMDNDLRRGKPTCHKAYDEATLADALQA 120  
Db 61 GLPENVLDPACAVAFIHYVSLIHDLPAMDNDLRRGKPTCHKAYDEATLADALQA 120  
QY 121 LAPEVLNDPGITVDAPARLKMITALTRASQGVGGQADLGSVGRKLTLPLENWHI 180  
Db 121 LAPEVLNDPGITVDAPARLKMITALTRASQGVGGQADLGSVGRKLTLPLENWHI 180  
QY 181 HKTGALIRASVNLAAALSKPDLDTCAVAKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240  
Db 181 HKTGALIRASVNLAAALSKPDLDTCAVAKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240  
QY 241 GKDINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRELSLYIERTH 297  
Db 241 GKDINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRELSLYIERTH 297

RESULT 3

US-09-941-947A-20  
; Sequence 20, Application US/09941947A  
; Publication No. US20030003528A1  
; GENERAL INFORMATION:  
; APPLICANT: Brzostowicz, Patricia C.  
; APPLICANT: Cheng, Qiong  
; APPLICANT: DiCosimo, Deana J.

; APPLICANT: Koffas, Mattheos  
; APPLICANT: Miller, Edward S. Jr.  
; APPLICANT: Odom, J. Martin  
; APPLICANT: Picataggio, Steve  
; APPLICANT: Rouviere, Pierre E.  
; TITLE OF INVENTION: CAROTENOID PRODUCTION FROM A SINGLE CARBON SOURCE  
; FILE REFERENCE: CL1903 US NA  
; CURRENT APPLICATION NUMBER: US/09/941,947A  
; CURRENT FILING DATE: 2001-09-01  
; PRIOR APPLICATION NUMBER: 60/229,907  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,858  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 60  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 20  
; LENGTH: 297  
; TYPE: PRT  
; ORGANISM: Methylobionas 16a  
US-09-941-947A-20

Query Match 100.0%; Score 1510; DB 10; Length 297;  
Best Local Similarity 100.0%; Pred. No. 4.4e-152;  
Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKLAAYLTVCQVERALDARLPAENILPQTLHQAMRYSVLNGGKTRPLTYATGOAL 60  
Db 1 MSKLAAYLTVCQVERALDARLPAENILPQTLHQAMRYSVLNGGKTRPLTYATGOAL 60  
QY 61 GLPENVLDPACAVAFIHYVSLIHDLPAMDNDLRRGKPTCHKAYDEATLADALQA 120  
Db 61 GLPENVLDPACAVAFIHYVSLIHDLPAMDNDLRRGKPTCHKAYDEATLADALQA 120  
QY 121 LAPEVLNDPGITVDAPARLKMITALTRASQGVGGQADLGSVGRKLTLPLENWHI 180  
Db 121 LAPEVLNDPGITVDAPARLKMITALTRASQGVGGQADLGSVGRKLTLPLENWHI 180  
QY 181 HKTGALIRASVNLAAALSKPDLDTCAVAKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240  
Db 181 HKTGALIRASVNLAAALSKPDLDTCAVAKLDHYAKICIGLSFQVKDDILDIEADTATLGKTQ 240  
QY 241 GKDINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRELSLYIERTH 297  
Db 241 GKDINDKPTYPALLGMAGAKQAKQELHEQAVESLTGFGSEADLLRELSLYIERTH 297

RESULT 4

US-09-815-242-11971  
; Sequence 11971, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA 011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
BEST FILING DATE: 2000-12-22  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: Fast-Seq for Windows Version 4.0  
SEQ ID NO 11971  
LENGTH: 295

TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
S-09-815-242-11971

Query Match 55.0%; Score 831; DB 9; Length 295;  
Best Local Similarity 60.3%; Pred. No. 8.6e-80;  
Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;

Y 4 LKAYLVCCERVERALDARLPANILPQTLHQAMRYSVLNGKRTPLLTATGQALGLP 63  
b 2 IAYQARQARVDAALDAFVAPRELOLYEAMRYSVWNGKRVPLLYAAACALGGA 61  
Y 64 ENVLDPACAVEPIHYVSLIHDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 123  
b 62 PQADAAACAVELIHAYSLVHDDLPAWDDDLRRGPTTHRAFDEATAILAGDALQALAF 121  
Y 124 EVLANDPGITVDAPARKMITALTTRASGQGMVGGQAIIDLSVGRKLTLPLENMHIHKT 183  
b 122 EVLADTRRNPEHVGLEMLTLRAAGSAGMVGGQAIIDLSVGVVALDQAALEVMHREKT 181  
Y 184 GALIRASVNLAAISKPDLOTCAKLDHYKACIGLSFQVKDDILDIEADTATLGKTKQKD 243  
b 182 GALIEASVRLGALASGRASPASLAALRYAEALGAFQVQDDILDVESETATLGKTKQKD 241  
Y 244 INDNKPYPALLGMACAKQKQAEHQAESLTGFGSEADLLRELSLYIER 295  
b 242 QAHNKPTYPALLGLEAKGVALELRDLAALDGPSPADPLQLARYIVER 293

## ESULT 5

S-10-369-493-8506  
Sequence 8506, Application US/10369493  
Publication No. US20030233675A1

## GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 8506

LENGTH: 291

TYPE: PRT

ORGANISM: Ralstonia metallidurans

S-10-369-493-8506

Query Match 52.4%; Score 791.5; DB 15; Length 291;  
Best Local Similarity 57.8%; Pred. No. 1.3e-75;  
Matches 163; Conservative 37; Mismatches 79; Indels 3; Gaps 2;

Y 14 RVREALDARLPANILPQTLHQAMRYSVLNGKRTPLLTATGQALGLPENVLDAPACA 73  
b 13 RTEAALALPSTDTTFYTLHEAMRYAVLGGKRVPLLVHAAEVEVGATPEACDAACA 72  
Y 74 VEPHYVSLIHDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVRLANDPGIT 133  
b 73 VENIHAYSLVHDDMPCCDDDLRRGRPTVHKAYDEATAILVGDALQSQAFVLAQTALA 132

QY 134 VDAPARKMITALTTRASGQGMVGGQAIIDLSVGRKLTLPLENMHIHKTGALIRASVNL 193  
DB 133 --REARKVVAELAVASGSGCMCGQAIIDLVNGKMTREALEGEMTKTGALLRSVVM 190  
QY 194 AALSKEPDLTCAVAKLDHYKACIGLSFQVKDDILDIEADTATLGKTKQKDINDKPTYP 253  
DB 191 GALTGQALVALDRAAAVGLAFQVQDDILDVTDATATLGKTAGKDAANDKPTYS 249  
QY 254 LGMAGAKQKQAEHQAESLTGFGSEADLLRELSLYIER 295  
DB 250 LGLDARELAALQRTDAHEALEGFGTRAGRLAELADLIVR 291

## RESULT 6

US-10-369-493-13999  
Sequence 13999, Application US/10369493  
Publication No. US20030233675A1

## GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 13999

LENGTH: 295

TYPE: PRT

ORGANISM: Pseudomonas fluorescens

US-10-369-493-13999

Query Match 51.8%; Score 782; DB 15; Length 295;

Best Local Similarity 56.9%; Pred. No. 1.4e-74;

Matches 168; Conservative 36; Mismatches 85; Indels 6; Gaps 4;

QY 4 LKAYLVCCERVERALDARLPANILPQ--TLHQAMRYSVLNGKRTPLLTATGQALG 61  
DB 2 IAYATSQARVVAALLETFLNAP--LPELARLYEAMRYSVWNGKRVPLLYAAACALG 59  
QY 62 LPENVLDAPACAVEPIHYVSLIHDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQAL 121  
DB 60 GKAEQANGAACAVELIHAYSLVHDDLPAWDDDLRRGPTTHKTFDEACAILAGDGLQSL 119  
QY 122 APEVLANDPGIT--VDAPARKMITALTTRASGQGMVGGQAIIDLSVGRKLTLPLENMHI 180  
DB 120 AFSALL--DPLSLDSADIRLQVVTALAHAGAGAGMVGGQAIIDLSVGLKLDQKALEQMR 178  
QY 181 HKTGALIRASVNLAAISKPDLTCAVAKLDHYKACIGLSFQVKDDILDIEADTATLGKTKQ 240  
DB 179 HKTGALIEVSVKLGALASGCAEKDELKSLQTYAALGAFQVQDDILDVESDETILGKTKQ 238  
QY 241 GKDDINDKPTYPALLGMACAKQKQAEHQAESLTGFGSEADLLRELSLYIER 295  
DB 239 GADIARDKPTYPALLGLDANKAYALELRDQALHRLRPFDAAEPLRLDLYIVR 293

## RESULT 7

US-10-369-493-4670  
Sequence 4670, Application US/10369493  
Publication No. US20030233675A1

## GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 FILE REFERENCE: 38-10(52052)B  
 CURRENT APPLICATION NUMBER: US/10369,493  
 PRIOR FILING DATE: 2003-02-28  
 PRIOR APPLICATION NUMBER: US 60/360,039  
 PRIOR FILING DATE: 2002-02-21  
 NUMBER OF SEQ ID NOS: 47374  
 SEQ ID NO 4670  
 LENGTH: 294  
 TYPE: PRT  
 ORGANISM: Burkholderia fungorum  
 US-10-369-493-4670

Query Match 47.6%; Score 719.5; DB 15; Length 294;  
 Best Local Similarity 53.2%; Pred. No. 6.3e-68;  
 Matches 151; Conservative 38; Mismatches 90; Indels 5; Gaps 2;  
 QY 13 ERVERALDARLPAENILPOTLHOAMRYSVLNGGKRTPLTYATGQALGLPENVLDPAC 72  
 DB 15 ERVETALDHYLPGATEPATLHEAMRYAVLGGKRVPLCHAGELTGARAECLDAAA 74  
 QY 73 AVEFIHYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGI 132  
 DB 75 ALEMHIYSLVHDDPMDDDALRRGKPTVHVKYDEATALLVGDALQALQAFVLTSD--- 131  
 QY 133 TVDAPA-RLKMITALTTRASGQGVGGQIDLGVSGRKLTLPLENHIIHKTGALIRASV 191  
 DB 132 -VLAFAQQAALVRELALASGVGCGQIDLASVGHITRTQLETWHRMKTGALLRAV 190  
 QY 192 NLAALSKEPDLDTCVAKKLDHYAKCIGLSFQVKDDILDTEADTATLTKTQKGDINDKPT 251  
 DB 191 RMGALAGETPDAAAMRSLDAYSAAVGLAFQVVDLIDVTTSATLGTAGDKAGDKPT 250  
 QY 252 PALLGNAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 295  
 DB 251 VSIIGLDASRALAAQLRSDAHLAPFGARAQRLAELADLVNR 294

RESULT 8  
 US-10-369-493-7429  
 Query Match 47.6%; Score 719.5; DB 15; Length 294;  
 Best Local Similarity 53.2%; Pred. No. 6.3e-68;  
 Matches 151; Conservative 38; Mismatches 90; Indels 5; Gaps 2;  
 QY 13 ERVERALDARLPAENILPOTLHOAMRYSVLNGGKRTPLTYATGQALGLPENVLDPAC 72  
 DB 15 ERVETALDHYLPGATEPATLHEAMRYAVLGGKRVPLCHAGELTGARAECLDAAA 74  
 QY 73 AVEFIHYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGI 132  
 DB 75 ALEMHIYSLVHDDPMDDDALRRGKPTVHVKYDEATALLVGDALQALQAFVLTSD--- 131  
 QY 133 TVDAPA-RLKMITALTTRASGQGVGGQIDLGVSGRKLTLPLENHIIHKTGALIRASV 191  
 DB 132 -VLAFAQQAALVRELALASGVGCGQIDLASVGHITRTQLETWHRMKTGALLRAV 190  
 QY 192 NLAALSKEPDLDTCVAKKLDHYAKCIGLSFQVKDDILDTEADTATLTKTQKGDINDKPT 251  
 DB 191 RMGALAGETPDAAAMRSLDAYSAAVGLAFQVVDLIDVTTSATLGTAGDKAGDKPT 250  
 QY 252 PALLGNAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 295  
 DB 251 VSIIGLDASRALAAQLRSDAHLAPFGARAQRLAELADLVNR 294

Query Match 47.6%; Score 719.5; DB 15; Length 294;  
 Best Local Similarity 53.2%; Pred. No. 6.3e-68;  
 Matches 151; Conservative 38; Mismatches 90; Indels 5; Gaps 2;  
 QY 13 ERVERALDARLPAENILPOTLHOAMRYSVLNGGKRTPLTYATGQALGLPENVLDPAC 72  
 DB 15 ERVETALDHYLPGATEPATLHEAMRYAVLGGKRVPLCHAGELTGARAECLDAAA 74  
 QY 73 AVEFIHYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGI 132  
 DB 75 ALEMHIYSLVHDDPMDDDALRRGKPTVHVKYDEATALLVGDALQALQAFVLTSD--- 131

Db 75 ALEMHIYSLVHDDPMDDDALRRGKPTVHVKYDEATALLVGDALQALQAFVLTSD--- 131  
 QY 133 TVDAPA-RLKMITALTTRASGQGVGGQIDLGVSGRKLTLPLENHIIHKTGALIRASV 191  
 DB 132 -VLAFAQQAALVRELALASGVGCGQIDLASVGHITRTQLETWHRMKTGALLRAV 190  
 QY 192 NLAALSKEPDLDTCVAKKLDHYAKCIGLSFQVKDDILDTEADTATLTKTQKGDINDKPT 251  
 DB 191 RMGALAGETPDAAAMRSLDAYSAAVGLAFQVVDLIDVTTSATLGTAGDKAGDKPT 250  
 QY 252 PALLGNAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 295  
 DB 251 VSIIGLDASRALAAQLRSDAHLAPFGARAQRLAELADLVNR 294

RESULT 9  
 US-10-369-493-15449  
 Query Match 47.4%; Score 715; DB 15; Length 291;  
 Best Local Similarity 53.2%; Pred. No. 1.9e-67;  
 Matches 151; Conservative 35; Mismatches 94; Indels 4; Gaps 2;  
 QY 14 RVERALDARLPAENILPOTLHOAMRYSVLNGGKRTPLTYATGQALGLPENVLDPAC 73  
 DB 12 RTERQDADCLPSPALAPQRLHMAHVAVLGGKRVPLLVYASGALFGAEGSLDAPAVA 71  
 QY 74 VEFIHYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGI 133  
 DB 72 VELIHAYSLVHDDLPAMDNDLRRGKPTVHVIAFDEATAILAGDALQARAFELLABAP--- 128  
 QY 134 VDAPALKMITALTTRASGQGVGGQIDLGVSGRKLTLPLENHIIHKTGALIRASVNL 193  
 DB 129 ADAALRVGWLQSLASAGAGCGGQALDIDATGQVQALQALQALAFVLANDPGI 188  
 QY 194 AALSKEPDLDTCVAKKLDHYAKCIGLSFQVKDDILDTEADTATLTKTQKGDINDKPT 253  
 DB 189 GVLT-GGAGIADQQLDAPADALGAFQVDDILDVSSSAQLGKTAGDKAAQSKSTYPA 247  
 QY 254 LIGMAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 297  
 DB 248 LIGMAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 291

Query Match 47.4%; Score 715; DB 15; Length 291;  
 Best Local Similarity 53.2%; Pred. No. 1.9e-67;  
 Matches 151; Conservative 35; Mismatches 94; Indels 4; Gaps 2;  
 QY 14 RVERALDARLPAENILPOTLHOAMRYSVLNGGKRTPLTYATGQALGLPENVLDPAC 73  
 DB 12 RTERQDADCLPSPALAPQRLHMAHVAVLGGKRVPLLVYASGALFGAEGSLDAPAVA 71  
 QY 74 VEFIHYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGI 133  
 DB 72 VELIHAYSLVHDDLPAMDNDLRRGKPTVHVIAFDEATAILAGDALQARAFELLABAP--- 128  
 QY 134 VDAPALKMITALTTRASGQGVGGQIDLGVSGRKLTLPLENHIIHKTGALIRASVNL 193  
 DB 129 ADAALRVGWLQSLASAGAGCGGQALDIDATGQVQALQALQALAFVLANDPGI 188  
 QY 194 AALSKEPDLDTCVAKKLDHYAKCIGLSFQVKDDILDTEADTATLTKTQKGDINDKPT 253  
 DB 189 GVLT-GGAGIADQQLDAPADALGAFQVDDILDVSSSAQLGKTAGDKAAQSKSTYPA 247  
 QY 254 LIGMAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 297  
 DB 248 LIGMAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 291

RESULT 10  
 US-10-369-493-15817  
 Query Match 47.4%; Score 715; DB 15; Length 291;  
 Best Local Similarity 53.2%; Pred. No. 1.9e-67;  
 Matches 151; Conservative 35; Mismatches 94; Indels 4; Gaps 2;  
 QY 14 RVERALDARLPAENILPOTLHOAMRYSVLNGGKRTPLTYATGQALGLPENVLDPAC 73  
 DB 12 RTERQDADCLPSPALAPQRLHMAHVAVLGGKRVPLLVYASGALFGAEGSLDAPAVA 71  
 QY 74 VEFIHYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGI 133  
 DB 72 VELIHAYSLVHDDLPAMDNDLRRGKPTVHVIAFDEATAILAGDALQARAFELLABAP--- 128  
 QY 134 VDAPALKMITALTTRASGQGVGGQIDLGVSGRKLTLPLENHIIHKTGALIRASVNL 193  
 DB 129 ADAALRVGWLQSLASAGAGCGGQALDIDATGQVQALQALQALAFVLANDPGI 188  
 QY 194 AALSKEPDLDTCVAKKLDHYAKCIGLSFQVKDDILDTEADTATLTKTQKGDINDKPT 253  
 DB 189 GVLT-GGAGIADQQLDAPADALGAFQVDDILDVSSSAQLGKTAGDKAAQSKSTYPA 247  
 QY 254 LIGMAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 297  
 DB 248 LIGMAGAKOKAQELHQAVESLTGFGSEADLLRELSLYIIR 291



APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 15817

LENGTH: 284

TYPE: PRT

ORGANISM: Xanthomonas campestris

S-10-369-493-15817

Query Match 46.8%; Score 707; DB 15; Length 284;  
Best Local Similarity 53.2%; Pred. No. 1.3e-66;  
Matches 150; Conservative 35; Mismatches 93; Indels 4; Gaps 2;

Y 14 RVERLDARLPAENILPQTLHQAQVSVLNGKTRPLLTATGQALGLENVLDAPACA 73  
b 7 RTERQLDACLPSALAPQRLHAAHRAVLGGKRMPLLVYASGALFGAEGSLDAPAVA 66  
Y 74 VEFHVSILHDDLPAMDNDLRRGKPTCHKAYDRAATILAGDALQALAFVLANDPGIT 133  
b 67 VELIHAYSLVHDDLPAMDNDLRRGKPTCHKAYDRAATILAGDALQALAFVLANDPGIT 123  
Y 134 VDAPARLKMITALTRASGQGMVGGQAIIDLSVGRKLTLPENLHHTKGTALIRASVNL 193  
b 124 ADAALRVGLQSLASAGAAGMCGGQALDIDATGQVQALEALQRMHAKTGTALIRAAVRM 183  
Y 194 AALSKPDLTCVAKKLDHYAKCIGLSFQVKDDILDIEDATILGKTQKQIDNDKPTTYP 253  
b 184 GVLT-GGAGIADQORLDADAFALGLAFQVRDDILDVSSAQLGKTAGKDAQAQSKSTYPA 242  
Y 254 LLGMAGAKQAKQELHQAQVESLTFGSGEADLLRELSTYIER 295  
b 243 LLGMDGAKTKLAELASHMHALLLPYGESGATLASIGRFAVDR 284

RESULT 11

S-10-369-493-16200

Sequence 16200, Application US/10369493

Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 16200

LENGTH: 284

TYPE: PRT

ORGANISM: Xanthomonas campestris

S-10-369-493-16200

Query Match 46.8%; Score 707; DB 15; Length 284;  
Best Local Similarity 53.2%; Pred. No. 1.3e-66;  
Matches 150; Conservative 35; Mismatches 93; Indels 4; Gaps 2;

Y 14 RVERLDARLPAENILPQTLHQAQVSVLNGKTRPLLTATGQALGLENVLDAPACA 73  
b 7 RTERQLDACLPSALAPQRLHAAHRAVLGGKRMPLLVYASGALFGAEGSLDAPAVA 66  
Y 74 VEFHVSILHDDLPAMDNDLRRGKPTCHKAYDRAATILAGDALQALAFVLANDPGIT 133

Db 67 VELIHAYSLVHDDLPAMDNDLRRGKPTCHKAYDRAATILAGDALQALAFVLANDPGIT 123  
QY 134 VDAPARLKMITALTRASGQGMVGGQAIIDLSVGRKLTLPENLHHTKGTALIRASVNL 193  
Db 124 ADAALRVGLQSLASAGAAGMCGGQALDIDATGQVQALEALQRMHAKTGTALIRAAVRM 183  
QY 194 AALSKPDLTCVAKKLDHYAKCIGLSFQVKDDILDIEDATILGKTQKQIDNDKPTTYP 253  
Db 184 GVLT-GGAGIADQORLDADAFALGLAFQVRDDILDVSSAQLGKTAGKDAQAQSKSTYPA 242  
QY 254 LLGMAGAKQAKQELHQAQVESLTFGSGEADLLRELSTYIER 295  
Db 243 LLGMDGAKTKLAELASHMHALLLPYGESGATLASIGRFAVDR 284

RESULT 12

US-09-815-242-14084

Sequence 14084, Application US/09815242

PATENT NO. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Olesen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-05-23

PRIOR FILING DATE: 2000-07-27

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-10-23

PRIOR FILING DATE: 2000-11-27

PRIOR FILING DATE: 2000-12-22

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 14084

LENGTH: 299

TYPE: PRT

ORGANISM: Salmonella typhi

US-09-815-242-14084

Query Match 46.7%; Score 705.5; DB 9; Length 299;  
Best Local Similarity 52.7%; Pred. No. 2e-66;  
Matches 156; Conservative 39; Mismatches 94; Indels 7; Gaps 3;

QY 3 KLKAYLTVCQERVARDLPAENILPQTLHQAQVSVLNGKTRPLLTATGQALGL 62  
Db 6 QLQACVQANQALSREI-APLPQN---TPVVEAMQYGALLSGKRLRPFLVYATGQMGV 61  
QY 63 PENYLDAPACAVEFHYVLSIHDDLPAMDNDLRRGKPTCHKAYDRAATILAGDALQALA 122  
Db 62 STATLDAPAAAVECHAYSLIHDDLPAMDNDLRRGKPTCHKAYDRAATILAGDALQALA 121  
QY 123 FEVLNDPGLTVDAPARLKMITALTRASGQGMVGGQAIIDLSVGRKLTLPENLHHTK 182  
Db 122 FTIISDAPMEVADRDRKIAMLASAGIAGCGQALDLAEGQRIITLDALERTHEK 181  
QY 183 TGAIRASVNLAAALSKPDLTCVAKKLDHYAKCIGLSFQVKDDILDIEDATILGKTQK 242

```
Db 182 TGAATRAAVRLGALSAGDKGNTLPILDRYAESIGLAFQVQDDILDVVGDTATLGKQGA 241
QY 243 DIDNDKPTYPALLGMAGAKQAELHEQAVESLTFGSEA---DLLRELSLYIER 295
Db 242 DQQLCKSTVPALLGLEQARNKARDLIEDARQSLHQLAAQSLDTSALEALANYIIOR 297

RESULT 13
US-09-815-242-10069
; Sequence 10069, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karli L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10069
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10069

Query Match 45.8%; Score 691.5; DB 9; Length 299;
Best Local Similarity 52.0%; Pred. No. 6.2e-65;
Matches 153; Conservative 36; Mismatches 96; Indels 9; Gaps 3;

QY 8 LTVQCVVERALD---ARLPAENILPQTLHQAMRYSVLNGGKRTPELLTYATGQALGPE 64
Db 7 LEACVKQANQALSRIAPLPFCN---TPVVTMQYVGLGKRLRPFLVYATGHRMGVST 63
QY 65 NVLDAPACAVPEFIHYVSLIHDDLPAMNDLRRGKPTCHKAYDEATAILAGDALQALAFE 124
Db 64 NTLDPAAAVECIHAYSLIHDDLPAMDDDLRRGLPTCHKVKGFEANAILAGDALQTLAFS 123
QY 125 VLANDPGITVDAPARKMTALTTRASGQGMVGQADLDGSGVGRKLTPELNNHHTKG 184
Db 124 ILSDADMEVSDRDEISMSLASASGIAGMCGQALDLDAGKHVPDLALERHHRKTG 183
QY 185 ALIRASVNLAKSPDLDTCAVKLDHYAKCIGLSFOVKDDILDIEDATATLGKTKQKDI 244
Db 184 ALIRAAVRLGALSAGDKGRALPVLDKYAESIGLAFQVQDDILDVVGDTATLGKQCADQ 243
QY 245 DNDKPTYPALLGMAGAKQAELHEQAVESLTFGSEA---DLLRELSLYIER 295
Db 244 QLGKSTYPALLGLEQARKKARDLIDDAQSLKQLAQSLDTSALEALADYIIOR 297

RESULT 14
US-10-369-493-732
; Sequence 732, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 732
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-369-493-732

Query Match 45.8%; Score 691.5; DB 15; Length 299;
Best Local Similarity 52.0%; Pred. No. 6.2e-65;
Matches 153; Conservative 36; Mismatches 96; Indels 9; Gaps 3;

QY 8 LTVQCVVERALD---ARLPAENILPQTLHQAMRYSVLNGGKRTPELLTYATGQALGPE 64
Db 7 LEACVKQANQALSRIAPLPFCN---TPVVTMQYVGLGKRLRPFLVYATGHRMGVST 63
QY 65 NVLDAPACAVPEFIHYVSLIHDDLPAMNDLRRGKPTCHKAYDEATAILAGDALQALAFE 124
Db 64 NTLDPAAAVECIHAYSLIHDDLPAMDDDLRRGLPTCHKVKGFEANAILAGDALQTLAFS 123
QY 125 VLANDPGITVDAPARKMTALTTRASGQGMVGQADLDGSGVGRKLTPELNNHHTKG 184
Db 124 ILSDADMEVSDRDEISMSLASASGIAGMCGQALDLDAGKHVPDLALERHHRKTG 183
QY 185 ALIRASVNLAKSPDLDTCAVKLDHYAKCIGLSFOVKDDILDIEDATATLGKTKQKDI 244
Db 184 ALIRAAVRLGALSAGDKGRALPVLDKYAESIGLAFQVQDDILDVVGDTATLGKQCADQ 243
QY 245 DNDKPTYPALLGMAGAKQAELHEQAVESLTFGSEA---DLLRELSLYIER 295
Db 244 QLGKSTYPALLGLEQARKKARDLIDDAQSLKQLAQSLDTSALEALADYIIOR 297

RESULT 15
US-10-369-493-10187
; Sequence 10187, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10187
; LENGTH: 306
; TYPE: PRT
; ORGANISM: magnetite-containing magnetic coccus
US-10-369-493-10187
```

Query Match 45.7%; Score 690; DB 15; Length 306;  
Best local Similarity 52.2%; Pred. No. 9.3e-65;  
Matches 157; Conservative 36; Mismatches 94; Indels 14; Gaps 5;

3 KUKAYLVUCQERVERALDARLPAPENILPQTLHQMRYSVLNGGRTRPLLTATGQALGL 62  
14 ELKPYLADRKYLIVESALDRLLPAPADREPRALTGMRYSLIGGRKLRPLTLACCEAVG- 72  
63 PENVLDP---ACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQ 119  
73 --GALERAMNFACAMECVHTYSLIHDDLPAMDNDLRRGHPTCHRYQGEATAILAGDALL 130  
120 ALAFEVLANDPGITVDAPARLKMITALTRASSQGWGQAIIDLGSVGRKLTLPLENNMH 179  
131 THAFE-LAAKPVTVGLPEARLELIGELAMAAAGTYGWWGQMLDIQGEKRPLELVQLQNIH 189  
180 IHKTGALIRASVNLAL---SKPDLDTCAVKKLDHYAKCIGLSFQVKDDILDIEADTATL 236  
190 IHKTGALIRVACLAGARLGGTTPD---QVHLKRYGEALGLAFQITDILLDEVGDTKL 245  
237 GKTQKGDINDKPTYPALIGMAGAKQAKELHEQAVESLTGFGSEADLLRELSLYIIERT 296  
246 GRATGVDRALDKATYPKLMGLAQAREAEENLMQEATRCLTTEFGQADFLRALAQYVIART 305  
297 H 297  
306 H 306

Search completed: February 29, 2004, 15:27:46  
Job time : 22.7191 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - protein search, using sw model

un on: February 29, 2004, 14:35:44 ; Search time 9.02528 Seconds  
(without alignments)  
1698.885 Million cell updates/sec

title: US-09-941-947A-20

effect score: 1510

sequence: 1 MSKLKAYLTCQERVERALD.....FGSEADLLRLSLYIERTH 297

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

Issued Patents AA:\*

1: /cgn2\_6/prodata2/iaa/5A\_COMB.pep:\*

2: /cgn2\_6/prodata2/iaa/53\_COMB.pep:\*

3: /cgn2\_6/prodata2/iaa/6A\_COMB.pep:\*

4: /cgn2\_6/prodata2/iaa/53\_COMB.pep:\*

5: /cgn2\_6/prodata2/iaa/PCUTS\_COMB.pep:\*

6: /cgn2\_6/prodata2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1510	100.0	297	4	US-09-934-903-14
2	831	55.0	401	4	US-09-252-991A-26387
3	702.5	46.5	307	4	US-09-343-681A-5508
4	693.5	45.9	299	4	US-09-489-039A-12733
5	605.5	40.1	308	4	US-09-328-352-4190
6	588.5	39.0	393	3	US-09-187-050-2
7	588.5	39.0	393	3	US-09-187-050-14
8	588.5	39.0	393	3	US-09-187-050-16
9	588.5	39.0	393	3	US-09-187-050-18
10	588.5	39.0	393	3	US-09-187-050-20
11	588.5	39.0	393	3	US-09-187-050-22
12	588.5	39.0	393	3	US-09-187-050-24
13	588.5	39.0	393	3	US-09-187-050-26
14	588.5	39.0	393	3	US-09-187-050-27
15	588.5	39.0	393	3	US-09-187-050-28
16	588.5	39.0	393	3	US-09-187-050-29
17	588.5	39.0	393	3	US-09-187-050-30
18	588.5	39.0	393	3	US-09-187-050-31
19	588.5	39.0	393	3	US-09-187-050-32
20	588.5	39.0	393	3	US-09-187-050-33
21	588.5	39.0	393	3	US-09-187-050-34
22	587	38.9	285	3	US-09-187-050-12
23	578	38.3	297	1	US-08-534-910B-7
24	577	38.2	297	3	US-08-586-466-2
25	577	38.2	297	3	US-09-475-304-2
26	577	38.2	297	4	US-09-101-126-3
27	577	38.2	297	4	US-09-367-528A-5

28	575	38.1	297	1	US-08-534-910B-10	Sequence 10, Appl
29	573	37.3	297	4	US-09-367-528A-1	Sequence 1, Appl
30	572	37.9	297	1	US-08-534-910B-6	Sequence 6, Appl
31	572	37.9	297	1	US-08-534-910B-9	Sequence 9, Appl
32	571	37.8	297	4	US-09-367-528A-3	Sequence 3, Appl
33	567	37.5	297	1	US-08-534-910B-8	Sequence 8, Appl
34	551	36.5	377	4	US-09-420-211-2	Sequence 2, Appl
35	529	35.0	302	4	US-09-134-001C-3171	Sequence 3171, Ap
36	512	33.9	342	4	US-09-540-236-2612	Sequence 2612, Ap
37	509.5	33.7	293	3	US-09-276-873-2	Sequence 2, Appl
38	505	33.4	280	4	US-09-634-238-303	Sequence 303, App
39	485.5	32.2	312	4	US-09-107-532A-6724	Sequence 6724, Ap
40	441.5	29.2	291	3	US-09-275-742-2	Sequence 2, Appl
41	336	22.3	302	1	US-07-783-705A-1	Sequence 1, Appl
42	318.5	21.1	293	3	US-08-980-832-37	Sequence 37, Appl
43	318.5	21.1	293	4	US-09-920-923B-37	Sequence 37, Appl
44	317.5	21.0	298	1	US-08-095-726-4	Sequence 4, Appl
45	317.5	21.0	298	1	US-08-096-043-4	Sequence 4, Appl

## ALIGNMENTS

### RESULT 1

US-09-934-903-14

; Sequence 14, Application US/09934903

; Patent No. 6660507

; GENERAL INFORMATION:

; APPLICANT: Koffas, Mattheos

; APPLICANT: Odum, J. Martin

; APPLICANT: Schenzle, Andreas J.

; APPLICANT: No. 6660507ton, Kelley C.

; APPLICANT: Tomb, Jean-Francois

; APPLICANT: Rouviere, Pierre

; APPLICANT: Picataggio, Stephen

; APPLICANT: Cheng, Qiong

; TITLE OF INVENTION: Genes Involved in Isoprenoid Compounds Production

; FILE REFERENCE: CUI646 US NA

; CURRENT APPLICATION NUMBER: US/09/934,903

; CURRENT FILING DATE: 2001-08-22

; PRIOR APPLICATION NUMBER: 60/229,907

; PRIOR FILING DATE: September 1, 2001

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 14

; LENGTH: 297

; TYPE: PRT

; ORGANISM: Methylomonas 16a

; FEATURE:

; OTHER INFORMATION: Amino acid sequences encoded by ORF7

; US-09-934-903-14

Query Match 100.0%; Score 1510; DB 4; Length 297;  
Best Local Similarity 100.0%; Pred. No. 8.6e-148;  
Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSKLKAYLTCQERVERALDARLPANILPOTLHQAMRYSVLNGGKRTPLLYATGOAL	60
Db	1	MSKLKAYLTCQERVERALDARLPANILPOTLHQAMRYSVLNGGKRTPLLYATGOAL	60
Qy	61	GLPENVLDPACAVEFHVYSLIHDDLPAWNNDDLARGKPTCHKAYDEATAILAGDALQA	120
Db	61	GLPENVLDPACAVEFHVYSLIHDDLPAWNNDDLARGKPTCHKAYDEATAILAGDALQA	120
Qy	121	LAPEVLNDPGITVDAPARKMTALTTRASGSGMVGGAIDLGSVGRKLTLPLENMHI	180
Db	121	LAPEVLNDPGITVDAPARKMTALTTRASGSGMVGGAIDLGSVGRKLTLPLENMHI	180
Qy	181	HKTGALLRASVNLAAASKPDLDTCVAKLDHYAKCIGLSFQVKDDILDIEADTATLGTKQ	240
Db	181	HKTGALLRASVNLAAASKPDLDTCVAKLDHYAKCIGLSFQVKDDILDIEADTATLGTKQ	240
Qy	241	GKOIDNDKPTYPALLGAGAKAKQAQLHEQAVESLTGFGSEADLLRLSLYIERTH	297

Db 241 GSDINDKPTYPALLGNAGAKQKAEHQAVESITGFGSEADLLRELSLYIERTH 297  
|||||

## RESULT 2

US-09-252-991A-26387  
; Sequence 26387, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26387  
; LENGTH: 401  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26387

Query Match 55.0%; Score 831; DB 4; Length 401;  
Best Local Similarity 60.3%; Pred. No. 2e-77;  
Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;  
QY 4 LKAYLTVCQVERALDARLPAENILPOTLHOAMRYSVLNGGKTRPLLTVATQALGLP 63  
DB 108 IAAVQARQARVDAALDALFVAPRELOQLYEAMRYSVNGGKVRPLLAYACEALGGA 167  
QY 64 ENVLDPACAFEFHVSILIHDDLPAMDNDLRRGKPTCHKAYDEATILAGDALQALAF 123  
DB 168 PORADAAACAVELIHAYSLVHDDLPAMDNDLRRGKPTCHKAYDEATILAGDALQALAF 227  
QY 124 EVLANDPGITVDAPARLKMITALTRASSGQMGVGOALDLSGVGRKLTLPLENMHIKT 183  
DB 228 EVLADTRNPQHAVCLEMTRLARAGSAGMGVGOALDLSGVGRKLTLPLENMHIKT 287  
QY 184 GALIRASVNLAAKSPDLTVCVAKKLDHYAKICIGLSFQVKDDILDIDTATLGKTQKGD 243  
DB 288 GALIEASVRLGALASGRAPASLAALRYAEALGAFQVQDDILDVSDTATLGKTQKGD 347  
QY 244 INDKPTYPALLGNAGAKQKAEHQAVESITGFGSEADLLRELSLYIER 295  
DB 348 QAHNKPTYPALLGNAGAKQKAEHQAVESITGFGSEADLLRELSLYIER 399  
|||||

## RESULT 3

US-09-543-681A-5908  
; Sequence 5908, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 5908  
; LENGTH: 307  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-5908

Query Match 46.5%; Score 702.5; DB 4; Length 307;  
Best Local Similarity 52.4%; Pred. No. 2.5e-64;

Matches 150; Conservative 36; Mismatches 97; Indels 3; Gaps 1;  
QY 13 EBERVALDARLPAENILPOTLHOAMRYSVLNGGKTRPLLTVATQALGLPENVLDPAC 72  
DB 20 QXVDEALNQALQTLFSPDMPLSQAMRYGALLGGKRLRFVYAVGEMENVVAVLNDVPA 79  
QY 73 AVEFTHVYSILIHDDLPAMDNDLRRGKPTCHKAYDEATILAGDALQALAEVLANDPGI 132  
DB 80 AIECIHAYSLIHDDLPAMDNDLRRGKPTCHIEFGEANAILAGDALQTLAEIILAKNAMP 139  
QY 133 TYDAPARLKMITALTRASSGQMGVGOALDLSGVGRKLTLPLENMHIKTGALIRASV 192  
DB 140 DVAIADRVAMTAEATASGLAGMCGQALDDADSKSIDLVALEKIHLKTKGALIRAAVR 199  
QY 193 LAALSKPDLTVCVAKKLDHYAKICIGLSFQVKDDILDIDTATLGKTQKGDINDKPTYP 252  
DB 200 LGALSAGQKQGHVLPALDKYAHISGLAFQVQDDILDVSTGTEGKRGSGQDEACKSTYP 259  
QY 253 ALLGMAGAKQKAEHQAVESLT---GFGSEADLLRELSLYIER 295  
DB 260 ALLGLAQKQKAEHQAVESLT---GFGSEADLLRELSLYIER 305  
|||||

## RESULT 4

US-09-489-039A-12733  
; Sequence 12733, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 12733  
; LENGTH: 299  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-12733

Query Match 45.9%; Score 693.5; DB 4; Length 299;  
Best Local Similarity 51.2%; Pred. No. 2e-63;  
Matches 149; Conservative 38; Mismatches 101; Indels 3; Gaps 1;  
QY 8 LTVCOERVALDARLPAENILPOTLHOAMRYSVLNGGKTRPLLTVATQALGLPENVL 67  
DB 7 LQACVEQANEARLRRFIAFPQFNTLVEAMHYGALLGGKRLRPFVATGEMFVCRITL 66  
QY 68 DAPACAVESFIHVYSILIHDDLPAMDNDLRRGKPTCHKAYDEATILAGDALQALAEVLA 127  
DB 67 DAPAAAVECIHAYSLMHDDLPAMDNDLRRGKPTCHIKFGEANAILAGDALQTLAFSILS 126  
QY 128 NDPGITVDAPARLKMITALTRASSGQMGVGOALDLSGVGRKLTLPLENMHIKTGALI 187  
DB 127 DAPWYDVDPDRLEAMVSLSAQAGVAGMCGQALDQAEQQVDLQALERIHRKTGALI 186  
QY 188 RASVNLAAKSPDLTVCVAKKLDHYAKICIGLSFQVKDDILDIDTATLGKTQKGDIND 247  
DB 187 RAAVEMGALSAGERGRAALPALDRYAENIGLAFQVQDDILDVVGDTATLGKRGQADQQLG 246  
QY 248 KPTYPALLGNAGAKQKAEHQAVESITGFGSEADLLRELSLYIER 295  
DB 247 KSTYPALLGLSQARKKAHDLIADARRSLDELAQSLDTSALALANYIIOR 297  
|||||

## RESULT 5

US-09-328-352-4190  
; Sequence 4190, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 4190  
LENGTH: 308  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
S-09-328-352-4190

Query Match 40.1%; Score 605.5; DB 4; Length 308;  
Best Local Similarity 47.0%; Pred. No. 2.6e-54;  
Matches 143; Conservative 50; Mismatches 100; Indels 11; Gaps 6;  
Y 1 MSKLKA-YITVCQERVERALDAPENILPOTLHOAMRYSVLGGKRPRLTYATGQA 59  
b 9 LSQVNDLVKQQRICQDLTTLAFTSI-PEELKSAVHEAVMLGGKVRPALCYATASL 67  
Y 60 LGLPENVLDA--PACAVEFHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDA 117  
b 68 QDNP-NFAAARAAVAVELIHCYSLAHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDI 126  
Y 118 LQALAEVLA---NDPGITVDAPARLKMITALTRASGSGQGMVGOAIDLSGVGRKLTP 173  
b 127 LQMAFEVLGSRFLPDGQGGTDAIVLQIQL--ATASKWVGQVLDLQAEAKQISQD 184  
Y 174 ELENHHTKGTALIRASVNLAAALSKEDLDTCAVAKLDHYAKCIGLSFQVKDDILDIEADT 233  
b 185 ELENHHTKGTALIRASVNLAAALSKEDLDTCAVAKLDHYAKCIGLSFQVKDDILDIEADT 244  
Y 234 ATLGKTCQKIDNDKTYTALLGMAGAKQAKQELHQAQVESLTGSGSEADLLRELSLYII 293  
b 245 ETLGKTAGKDEQVOKSTYPALMGLAQIYAKELHQAQVESLTGSGSEADLLRELSLYII 304  
Y 294 ERTH 297  
b 305 ARTN 308

RESULT 6  
S-09-187-050-2  
Sequence 2, Application US/09187050B  
Patent No. 6043072  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney B  
APPLICANT: Hefner, Jerry  
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate  
FILE REFERENCE: WSUR12423  
CURRENT APPLICATION NUMBER: US/09/187,050B  
CURRENT FILING DATE: 1998-11-05  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 2  
LENGTH: 393  
TYPE: PRT  
ORGANISM: Taxus canadensis  
S-09-187-050-2

Query Match 39.0%; Score 588.5; DB 3; Length 393;  
Best Local Similarity 46.6%; Pred. No. 2.2e-52;  
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;  
Y 15 VERALDARLPAENILPOTLHOAMRYSVLGGKRPRLTYATGQALGPENVLDPACAV 74  
b 112 VDAALDKAIPLE--YPEKHESMRYSLLAGGRKVRPALCIAACELVGGSDLPMTACAM 169  
Y 75 EFHIVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAEVFLANDPGITV 134  
b 170 EMHTWSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAEVFLANDPGITV 229

Y 135 DAPARLKMITALTRASGSGQGMVGOAIDLSGVG-RKLTLPENLHHTKGTALIRASVNL 193  
b 230 PSDRTLREIVSELGKTTIGSQGLGVGVVDITSEGDANVDLKTLEWHIHKTAVALLECSVVS 289  
Y 194 AALSKPDLDTCVAKLDHYAKCIGLSFQVKDDILDIEADTATLGKTCQKIDNDKTYPA 253  
b 290 GGILGGATEDEIA-RIRRYARCVGLLFQVVDILDVTKSEELGKTAGKDLLTDKATYPK 348  
Y 254 LLGMAGAKQAKQELHQAQVESLTGFGS-EADLLRELSLYIIER 295  
b 349 LMGLEKAKPAEALATRAKEELSSFDQIKAAAPLLGLADYIAFR 391  
RESULT 7  
US-09-187-050-14  
Sequence 14, Application US/09187050B  
Patent No. 6043072  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney B  
APPLICANT: Hefner, Jerry  
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate  
FILE REFERENCE: WSUR12423  
CURRENT APPLICATION NUMBER: US/09/187,050B  
CURRENT FILING DATE: 1998-11-05  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: Patent in Ver. 2.0  
SEQ ID NO 14  
LENGTH: 393  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP  
OTHER INFORMATION: synthase protein variant  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (1)..(393)  
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein  
OTHER INFORMATION: variant  
US-09-187-050-14

Query Match 39.0%; Score 588.5; DB 3; Length 393;  
Best Local Similarity 46.6%; Pred. No. 2.2e-52;  
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;  
Y 15 VERALDARLPAENILPOTLHOAMRYSVLGGKRPRLTYATGQALGPENVLDPACAV 74  
b 112 VDAALDKAIPLE--YPEKHESMRYSLLAGGRKVRPALCIAACELVGGSDLPMTACAM 169  
Y 75 EFHIVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAEVFLANDPGITV 134  
b 170 EMHTWSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAEVFLANDPGITV 229  
Y 135 DAPARLKMITALTRASGSGQGMVGOAIDLSGVG-RKLTLPENLHHTKGTALIRASVNL 193  
b 230 PSDRTLREIVSELGKTTIGSQGLGVGVVDITSEGDANVDLKTLEWHIHKTAVALLECSVVS 289  
Y 194 AALSKPDLDTCVAKLDHYAKCIGLSFQVKDDILDIEADTATLGKTCQKIDNDKTYPA 253  
b 290 GGILGGATEDEIA-RIRRYARCVGLLFQVVDILDVTKSEELGKTAGKDLLTDKATYPK 348  
Y 254 LLGMAGAKQAKQELHQAQVESLTGFGS-BADLLRELSLYIIER 295  
b 349 LMGLEKAKPAEALATRAKEELSSFDQIKAAAPLLGLADYIAFR 391

RESULT 8  
US-09-187-050-16  
Sequence 16, Application US/09187050B  
Patent No. 6043072  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney B

APPLICANT: Hefner, Jerry  
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate  
TITLE OF INVENTION: Synthesis, And Methods of Use  
FILE REFERENCE: WSUR12423  
CURRENT APPLICATION NUMBER: US/09/187,050B  
CURRENT FILING DATE: 1998-11-05  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 16  
LENGTH: 393  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP  
OTHER INFORMATION: synthase protein variant  
NAME/KEY: VARIANT  
LOCATION: (1)...(393)  
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein  
OTHER INFORMATION: variant  
US-09-187-050-16

Query Match 39.0%; Score 588.5; DB 3; Length 393;  
Best Local Similarity 46.6%; Pred. No. 2.2e-52;  
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;  
15 VERALDARLPAENILPOTLHOAMRYSVLNGGKTRPLITYATQALGLPENVLDPACAV 74  
112 VDAALDKAIPLE--YPEKIHESMRYSLLAGKRVFPALCIAACELVGGSDLAMPTACAM 169  
75 EPIHVSILIHDDLPAAMNDLRRGKPTCHKAYDEATAILAGDALQALAEVLANDPGTV 134  
170 EMHTMSLIHDDLPCMDNDFRGGKPTNHKVFEDTAVLAGDALLSFAFEHIAVATSKTV 229  
135 DAPARLKMITALTTRASSQGMVGGQALDLSVG-RKLTPELENMHIHKTGALIRASVNL 193  
230 PSDRTLRLVISELKGITIGSGLVGGQVVDITSEGDANVDLTKLEWIIHKTAVLLECSVVS 289  
194 AALSKPDLDTVCVKKLHAKYACIGLSFQVXDDIILDIEADTATGKTGGKIDNDKPTYP 253  
290 GGILGGATEDEIA-RIRRYARCVELLFQVVDIILDVTKSSEELGKTAGKDLLTDKATY 348  
254 LEMGAKAKQAEALHEQAVESLTGFGS-EADLLRELSLYIIER 295  
349 LMGLEKAKEFAELATRAKEBELSSFDQIKAAPLLGLADYIAFR 391

RESULT 9  
US-09-187-050-18  
Sequence 18, Application US/09187050B  
Patent No. 6043072  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney B  
APPLICANT: Hefner, Jerry  
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate  
TITLE OF INVENTION: Synthesis, And Methods of Use  
FILE REFERENCE: WSUR12423  
CURRENT APPLICATION NUMBER: US/09/187,050B  
CURRENT FILING DATE: 1998-11-05  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 18  
LENGTH: 393  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP  
OTHER INFORMATION: synthase protein variant  
NAME/KEY: VARIANT  
LOCATION: (1)...(393)  
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein  
OTHER INFORMATION: variant

US-09-187-050-18

Query Match 39.0%; Score 588.5; DB 3; Length 393;  
Best Local Similarity 46.6%; Pred. No. 2.2e-52;  
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;  
15 VERALDARLPAENILPOTLHOAMRYSVLNGGKTRPLITYATQALGLPENVLDPACAV 74  
112 VDAALDKAIPLE--YPEKIHESMRYSLLAGKRVFPALCIAACELVGGSDLAMPTACAM 169  
75 EPIHVSILIHDDLPAAMNDLRRGKPTCHKAYDEATAILAGDALQALAEVLANDPGTV 134  
170 EMHTMSLIHDDLPCMDNDFRGGKPTNHKVFEDTAVLAGDALLSFAFEHIAVATSKTV 229  
135 DAPARLKMITALTTRASSQGMVGGQALDLSVG-RKLTPELENMHIHKTGALIRASVNL 193  
230 PSDRTLRLVISELKGITIGSGLVGGQVVDITSEGDANVDLTKLEWIIHKTAVLLECSVVS 289  
194 AALSKPDLDTVCVKKLHAKYACIGLSFQVXDDIILDIEADTATGKTGGKIDNDKPTYP 253  
290 GGILGGATEDEIA-RIRRYARCVELLFQVVDIILDVTKSSEELGKTAGKDLLTDKATY 348  
254 LEMGAKAKQAEALHEQAVESLTGFGS-EADLLRELSLYIIER 295  
349 LMGLEKAKEFAELATRAKEBELSSFDQIKAAPLLGLADYIAFR 391

RESULT 10

US-09-187-050-20  
Sequence 20, Application US/09187050B  
Patent No. 6043072  
GENERAL INFORMATION:  
APPLICANT: Croteau, Rodney B  
APPLICANT: Hefner, Jerry  
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate  
TITLE OF INVENTION: Synthesis, And Methods of Use  
FILE REFERENCE: WSUR12423  
CURRENT APPLICATION NUMBER: US/09/187,050B  
CURRENT FILING DATE: 1998-11-05  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 20  
LENGTH: 393  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP  
OTHER INFORMATION: synthase protein variant  
NAME/KEY: VARIANT  
LOCATION: (1)...(393)  
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein  
OTHER INFORMATION: variant  
US-09-187-050-20

Query Match 39.0%; Score 588.5; DB 3; Length 393;  
Best Local Similarity 46.6%; Pred. No. 2.2e-52;  
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;  
15 VERALDARLPAENILPOTLHOAMRYSVLNGGKTRPLITYATQALGLPENVLDPACAV 74  
112 VDAALDKAIPLE--YPEKIHESMRYSLLAGKRVFPALCIAACELVGGSDLAMPTACAM 169  
75 EPIHVSILIHDDLPAAMNDLRRGKPTCHKAYDEATAILAGDALQALAEVLANDPGTV 134  
170 EMHTMSLIHDDLPCMDNDFRGGKPTNHKVFEDTAVLAGDALLSFAFEHIAVATSKTV 229  
135 DAPARLKMITALTTRASSQGMVGGQALDLSVG-RKLTPELENMHIHKTGALIRASVNL 193  
230 PSDRTLRLVISELKGITIGSGLVGGQVVDITSEGDANVDLTKLEWIIHKTAVLLECSVVS 289  
194 AALSKPDLDTVCVKKLHAKYACIGLSFQVXDDIILDIEADTATGKTGGKIDNDKPTYP 253

```
290 GGIIGGATEDEIA-RIRRYARCVGLLFQVDDILDVTKSSBELGKTAGKOLLTDKATYPK 348
254 LIGMAGAKQAKQELHEQAVESLTGFGS-EADLLRSLSYIIER 295
349 LMGLEKAKEFAAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391

RESULT 11
3-09-187-050-22
Sequence 22, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Heifer, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 22
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
NAME/KEY: VARIANT
LOCATION: (1)..(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-22

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;

15 VERALDARLPAENILPQTLHQAMRYSVLNGGKRTPLITYATGQALGPENVLDPACAV 74
112 VDAALDKAIPLE--YPEKHESMRYSLLAGGKRVPAICIAACELVGGSDLPMTACAM 169
75 EPIHYVSLIHDDLPAMDNDLLRRGKPTCKAYDEATAILAGDALQALAFVLANDPGIV 134
170 EMHTWSLIHDDLPAMDNDLLRRGKPTCKAYDEATAILAGDALQALAFVLANDPGIV 229
135 DAPARKMITALTTRASGSGQMGVGGQAIIDLSVG-RKLTLPLENNHHIKTGALLIRASVNL 193
230 PSDRTLRLVISELKGITGSGQLVGGQWVDITSEGDANVDLKTLEWHIHKTAVLLECSVWS 289
194 AALSKPDLDTCAKKLDHYAKCIGLSFQVDDILDIEADTATLGKTQKQIDNDKPTYP 253
290 GGIIGGATEDEIA-RIRRYARCVGLLFQVDDILDVTKSSBELGKTAGKOLLTDKATYPK 348

254 LIGMAGAKQAKQELHEQAVESLTGFGS-EADLLRSLSYIIER 295
349 LMGLEKAKEFAAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391

RESULT 12
3-09-187-050-24
Sequence 24, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Heifer, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34

290 GGIIGGATEDEIA-RIRRYARCVGLLFQVDDILDVTKSSBELGKTAGKOLLTDKATYPK 348
254 LIGMAGAKQAKQELHEQAVESLTGFGS-EADLLRSLSYIIER 295
349 LMGLEKAKEFAAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391
```

```
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 24
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
NAME/KEY: VARIANT
LOCATION: (1)..(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-24

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;

15 VERALDARLPAENILPQTLHQAMRYSVLNGGKRTPLITYATGQALGPENVLDPACAV 74
112 VDAALDKAIPLE--YPEKHESMRYSLLAGGKRVPAICIAACELVGGSDLPMTACAM 169
75 EPIHYVSLIHDDLPAMDNDLLRRGKPTCKAYDEATAILAGDALQALAFVLANDPGIV 134
170 EMHTWSLIHDDLPAMDNDLLRRGKPTCKAYDEATAILAGDALQALAFVLANDPGIV 229
135 DAPARKMITALTTRASGSGQMGVGGQAIIDLSVG-RKLTLPLENNHHIKTGALLIRASVNL 193
230 PSDRTLRLVISELKGITGSGQLVGGQWVDITSEGDANVDLKTLEWHIHKTAVLLECSVWS 289
194 AALSKPDLDTCAKKLDHYAKCIGLSFQVDDILDIEADTATLGKTQKQIDNDKPTYP 253
290 GGIIGGATEDEIA-RIRRYARCVGLLFQVDDILDVTKSSBELGKTAGKOLLTDKATYPK 348

254 LIGMAGAKQAKQELHEQAVESLTGFGS-EADLLRSLSYIIER 295
349 LMGLEKAKEFAAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391

RESULT 13
US-09-187-050-26
Sequence 26, Application US/09187050B
Patent No. 6043072
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney B
APPLICANT: Heifer, Jerry
TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate
FILE REFERENCE: WSUR12423
CURRENT APPLICATION NUMBER: US/09/187,050B
CURRENT FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 393
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP
OTHER INFORMATION: synthase protein variant
NAME/KEY: VARIANT
LOCATION: (1)..(393)
OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein
OTHER INFORMATION: variant
US-09-187-050-26

Query Match 39.0%; Score 588.5; DB 3; Length 393;
Best Local Similarity 46.6%; Pred. No. 2.2e-52;
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;

15 VERALDARLPAENILPQTLHQAMRYSVLNGGKRTPLITYATGQALGPENVLDPACAV 74
112 VDAALDKAIPLE--YPEKHESMRYSLLAGGKRVPAICIAACELVGGSDLPMTACAM 169
75 EPIHYVSLIHDDLPAMDNDLLRRGKPTCKAYDEATAILAGDALQALAFVLANDPGIV 134
170 EMHTWSLIHDDLPAMDNDLLRRGKPTCKAYDEATAILAGDALQALAFVLANDPGIV 229
135 DAPARKMITALTTRASGSGQMGVGGQAIIDLSVG-RKLTLPLENNHHIKTGALLIRASVNL 193
230 PSDRTLRLVISELKGITGSGQLVGGQWVDITSEGDANVDLKTLEWHIHKTAVLLECSVWS 289
194 AALSKPDLDTCAKKLDHYAKCIGLSFQVDDILDIEADTATLGKTQKQIDNDKPTYP 253
290 GGIIGGATEDEIA-RIRRYARCVGLLFQVDDILDVTKSSBELGKTAGKOLLTDKATYPK 348

254 LIGMAGAKQAKQELHEQAVESLTGFGS-EADLLRSLSYIIER 295
349 LMGLEKAKEFAAEALATRAKEELSSFDQIKAAPLLGLADYIAFR 391
```



Db 112 VDAALDKAIPLE--YPEKIHESMRYSLLAGGKVRPALCIAACELVGGSDLAMPTACAM 169  
QY 75 EFHIVYSLIHDDLPAKDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 134  
Db 170 EMHTMSLIHDDLPCMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 229  
QY 135 DAPARKMTALTTRASGGQGVGGQAILDLSVG-RKLTPELENMHIHKTGALIRASVNL 193  
Db 230 PSDRTLRLVISELGTIGSQGLVGGQVVDITSEGDANVDLKTLEWIIHKTAVLECSVVS 289  
QY 194 AALSKPDLDTCVAKKLDHYAKCIGLSFQVDDILDIEADTATLGKTGKIDNDKPTYP 253  
Db 290 GGIIGGATEDEIA-RIRRYARCVGLLFGVDDILDVTSSSELGKTAGKOLLTDKATY 348  
QY 254 LLMGAKAKOQAEHLHEQAVESLTGFGS-EADLLRELSLYIER 295  
Db 349 LMGLEKAKEFAAEALATRAKEELSSPDQIKAAAPLLGLADYIAFR 391

RESULT 14  
US-09-187-050-27  
; Sequence 27, Application US/09187050B  
; Patent No. 6043072  
; GENERAL INFORMATION:  
; APPLICANT: Croteau, Rodney B  
; APPLICANT: Heiner, Jerry  
; TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate  
; TITLE OF INVENTION: Synthase, And Methods of Use  
; FILE REFERENCE: WSUR12423  
; CURRENT APPLICATION NUMBER: US/09/187,050B  
; CURRENT FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 27  
; LENGTH: 393  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP  
; OTHER INFORMATION: synthase protein variant  
; NAME/KEY: VARIANT  
; LOCATION: (1)..(393)  
; OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein  
; OTHER INFORMATION: variant  
US-09-187-050-27

Query Match 39.0%; Score 588.5; DB 3; Length 393;  
Best Local Similarity 46.6%; Pred. No. 2.2e-52;  
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;  
QY 15 VERALDARLPAENILPOTLHOAMRYSVLNGGKTRPPLTYATGQALGPLPENVDAPACAV 74  
Db 112 VDAALDKAIPLE--YPEKIHESMRYSLLAGGKVRPALCIAACELVGGSDLAMPTACAM 169  
QY 75 EFHIVYSLIHDDLPAKDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 134  
Db 170 EMHTMSLIHDDLPCMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 229  
QY 135 DAPARKMTALTTRASGGQGVGGQAILDLSVG-RKLTPELENMHIHKTGALIRASVNL 193  
Db 230 PSDRTLRLVISELGTIGSQGLVGGQVVDITSEGDANVDLKTLEWIIHKTAVLECSVVS 289  
QY 194 AALSKPDLDTCVAKKLDHYAKCIGLSFQVDDILDIEADTATLGKTGKIDNDKPTYP 253  
Db 290 GGIIGGATEDEIA-RIRRYARCVGLLFGVDDILDVTSSSELGKTAGKOLLTDKATY 348  
QY 254 LLMGAKAKOQAEHLHEQAVESLTGFGS-EADLLRELSLYIER 295  
Db 349 LMGLEKAKEFAAEALATRAKEELSSPDQIKAAAPLLGLADYIAFR 391

RESULT 15  
US-09-187-050-28  
; Sequence 28, Application US/09187050B  
; Patent No. 6043072  
; GENERAL INFORMATION:  
; APPLICANT: Croteau, Rodney B  
; APPLICANT: Heiner, Jerry  
; TITLE OF INVENTION: Nucleic Acids Encoding Taxus Geranylgeranyl Diphosphate  
; TITLE OF INVENTION: Synthase, And Methods of Use  
; FILE REFERENCE: WSUR12423  
; CURRENT APPLICATION NUMBER: US/09/187,050B  
; CURRENT FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 28  
; LENGTH: 393  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Taxus GGPP  
; OTHER INFORMATION: synthase protein variant  
; NAME/KEY: VARIANT  
; LOCATION: (1)..(393)  
; OTHER INFORMATION: Computer-generated Taxus GGPP synthase protein  
; OTHER INFORMATION: variant  
US-09-187-050-28

Query Match 39.0%; Score 588.5; DB 3; Length 393;  
Best Local Similarity 46.6%; Pred. No. 2.2e-52;  
Matches 132; Conservative 45; Mismatches 101; Indels 5; Gaps 4;  
QY 15 VERALDARLPAENILPOTLHOAMRYSVLNGGKTRPPLTYATGQALGPLPENVDAPACAV 74  
Db 112 VDAALDKAIPLE--YPEKIHESMRYSLLAGGKVRPALCIAACELVGGSDLAMPTACAM 169  
QY 75 EFHIVYSLIHDDLPAKDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 134  
Db 170 EMHTMSLIHDDLPCMDNDLRRGKPTCHKAYDEATAILAGDALQALAFEVLANDPGITV 229  
QY 135 DAPARKMTALTTRASGGQGVGGQAILDLSVG-RKLTPELENMHIHKTGALIRASVNL 193  
Db 230 PSDRTLRLVISELGTIGSQGLVGGQVVDITSEGDANVDLKTLEWIIHKTAVLECSVVS 289  
QY 194 AALSKPDLDTCVAKKLDHYAKCIGLSFQVDDILDIEADTATLGKTGKIDNDKPTYP 253  
Db 290 GGIIGGATEDEIA-RIRRYARCVGLLFGVDDILDVTSSSELGKTAGKOLLTDKATY 348  
QY 254 LLMGAKAKOQAEHLHEQAVESLTGFGS-EADLLRELSLYIER 295  
Db 349 LMGLEKAKEFAAEALATRAKEELSSPDQIKAAAPLLGLADYIAFR 391

Search completed: February 29, 2004, 14:54:48  
Job time : 11.0253 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

un on: February 29, 2004, 14:26:38 ; Search time 35.1152 Seconds  
(without alignments)  
2389.754 Million cell updates/sec

title: US-09-941-947A-20

effect score: 1510

sequence: 1 MSLKXAYLTVCCOERVERALD.....FGSEADLLRLSLVIERH 297

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

sarched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : A\_Geneseq\_29Jan04:\*

1: Geneseqp1980s:\*

2: Geneseqp1990s:\*

3: Geneseqp2000s:\*

4: Geneseqp2001s:\*

5: Geneseqp2002s:\*

6: Geneseqp2003as:\*

7: Geneseqp2003bs:\*

8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

result No.	Score	Match	Length	DB ID	Description
1	1510	100.0	297	5	Aae22308 Methylomo
2	1510	100.0	297	5	Abg61586 High grow
3	1510	100.0	297	5	Aau80331 Methylomo
4	831	55.0	295	4	Aau36378 Pseudomon
5	831	55.0	295	6	Abu38665 Protein e
6	811	53.7	295	6	Abu40141 Protein e
7	791	52.4	295	6	Abu41505 Protein e
8	733	48.5	298	6	Abp78160 N. gonorr
9	733	48.5	298	6	Abp80110 N. gonorr
10	733	48.5	298	6	Abu37216 Protein e
11	729	48.3	298	6	Abu38181 Protein e
12	719.5	47.6	293	6	Abu21390 Protein e
13	711	47.1	294	6	Abu19741 Protein e
14	711	47.1	299	6	Abu28056 Protein e
15	705.5	46.7	299	4	Aau38491 Salmonell
16	705.5	46.7	299	6	Abu47496 Protein e
17	705.5	46.7	310	6	Abm67133 Photorhab
18	704	46.6	310	6	Abu49248 Protein e
19	702.5	46.5	306	6	Abu50231 Protein e
20	702.5	46.5	307	6	Abu40769 Protein e
21	699	46.3	294	6	Abu22567 Protein e
22	693.5	45.9	299	6	Abu31979 Protein e
23	691.5	45.8	299	2	Aay52835 Escherich
24	691.5	45.8	299	4	Aau34476 E. coli c
25	691.5	45.8	299	5	Abu05280 Prenyl al

26	691.5	45.8	299	5	Aao22194 Protein r
27	691.5	45.8	299	5	Abg68085 E.coli pr
28	691.5	45.8	299	6	Abu28527 Protein e
29	691.5	45.8	299	7	Ade86721 Geranyltr
30	691	45.8	303	6	Abu23165 Protein e
31	690	45.7	295	6	Abu39094 Protein e
32	689.5	45.7	299	7	Ad886723 Geranyltr
33	683.5	45.3	299	5	Abu05283 Prenyl al
34	683.5	45.3	299	5	Abg68093 E.coli pr
35	682.5	45.2	299	5	Abu05282 Prenyl al
36	682.5	45.2	299	5	Abg68092 E.coli pr
37	681.5	45.1	299	5	Abu05281 E.coli f
38	681.5	45.1	299	5	Abg68091 Haemophil
39	677	44.8	295	4	Aau35646 Protein e
40	677	44.8	295	6	Abu30591 Protein e
41	675	44.7	291	7	Ad874157 X. fastid
42	610	40.4	369	5	Aae25932 Soybean G
43	610	40.4	369	6	Abu67329 Soybean G
44	605.5	40.1	308	6	Ada32903 Acinetoba
45	604	40.0	295	3	Aag42844 Arabidops

## ALIGNMENTS

### RESULT 1

Aae22308

ID AAE22308 standard; protein; 297 AA.

XX AC AAE22308;

XX 07-AUG-2003 (revised)

DT 25-JUL-2002 (first entry)

XX XX

DE Methylomonas 16a sp. farnesyl diphosphate synthetase (fspa) enzyme.

XX Carotenoid; isopentenyl pyrophosphate; antheraxanthin; astaxanthin; diet;

KW anti-oxidant; steroid; flavour; fragrance; electro-optic application;

KW aquaculture; enzyme; farnesyl diphosphate synthetase; fspa.

XX OS

XX Methylomonas sp.

XX FN

XX WO200218617-A2.

XX PD

XX 07-MAR-2002.

XX PF

XX 04-SEP-2001; 2001WO-US027420.

XX PR

XX 01-SEP-2000; 2000US-0229858P.

XX PA

XX 01-SEP-2000; 2000US-0229907P.

XX XX

XX (DUPO ) DU FONT DE NEMOURS & CO S I.

XX PI

XX Brzostowicz PC, Cheng Q, Dicosimo DJ, Koffas M, Miller ES;

XX PT

XX Odum JM, Picataggio SK, Rouviere PE;

XX DR

XX WPI; 2002-351711/38.

XX DR

XX N-PSDB; AAD35506.

XX PT

XX Producing carotenoid compounds e.g. antheraxanthin and astaxanthin, by

XX using microorganisms having a nucleic acid molecule encoding enzymes in

XX the carotenoid biosynthetic pathway and which metabolize single carbon

XX PT

XX substrates.

XX PS

XX Claim 35; Page 125-126; 156pp; English.

XX CC

XX The invention relates to a method for producing carotenoid compounds. The

XX method comprises a transformed metabolising host cell, comprising

XX suitable levels of isopentenyl pyrophosphate and a nucleic acid molecule

XX encoding an enzyme in the carotenoid biosynthetic pathway, under the

XX control of regulatory sequences, and contacting the host cell with carbon

XX substrate to produce a carotenoid compound. The method is useful for

XX producing carotenoid compounds such as antheraxanthin and astaxanthin, by

CC using microorganism having a nucleic acid molecule encoding enzymes in  
 CC the carotenoid biosynthetic pathway and which metabolise single carbon  
 CC substrates. The carotenoids have potent anti-oxidant properties useful in  
 CC diet, and aquaculture elements. The carotenoids are also useful as  
 CC intermediates in the synthesis of steroids flavours and fragrances and  
 CC compounds for potential electro-optic applications. The present sequence  
 CC is Methylomonas 16a sp. geranyltransferase or farnesyl diphosphate  
 CC synthetase (ispA) enzyme used in the invention. (Updated on 07-AUG-2003  
 CC to correct OS field.)  
 XX  
 SQ Sequence 297 AA;

Query Match 100.0%; Score 1510; DB 5; Length 297;  
 Best Local Similarity 100.0%; Pred. No. 7.3e-140;  
 Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKIKAYLVCCQERVERALDARLPAENILPQTLHOAMEYSVLNGGKTRPLLTATGQAL 60  
 DB 1 MSKIKAYLVCCQERVERALDARLPAENILPQTLHOAMEYSVLNGGKTRPLLTATGQAL 60  
 QY 61 GLPENVLDAPACAVEFIHVYSLIHDDLPAMNDLRRGKPTCHKAYDEATATLQAL 120  
 DB 61 GLPENVLDAPACAVEFIHVYSLIHDDLPAMNDLRRGKPTCHKAYDEATATLQAL 120  
 QY 121 LAPEVLNDPGITVDAPARLKMITALTRASGQGWGGQADLGSVGRKLTLPLENMHI 180  
 DB 121 LAPEVLNDPGITVDAPARLKMITALTRASGQGWGGQADLGSVGRKLTLPLENMHI 180  
 QY 181 HTGALIRASVNLAAALSKPDLDTCAKLDHVAKCIGLSFQVKDILDIADTATLQKTQ 240  
 DB 181 HTGALIRASVNLAAALSKPDLDTCAKLDHVAKCIGLSFQVKDILDIADTATLQKTQ 240  
 QY 241 GKIDNDKPTYPALLGMAGAKQAQELHEQAVESLTGFGSEADLLRELSLYIERTH 297  
 DB 241 GKIDNDKPTYPALLGMAGAKQAQELHEQAVESLTGFGSEADLLRELSLYIERTH 297

RESULT 2  
 ABG61586  
 ID ABG61586 standard; protein; 297 AA.

XX ABG61586;  
 AC  
 XX  
 XX  
 DT 07-AUG-2003 (revised)  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE High growth methanotrophic bacterial strain polypeptide #36.  
 XX  
 KW High growth methanotrophic bacterial strain; C1 carbon substrate; enzyme;  
 KW methane; methanol; Embden-Meyerhof carbon flux pathway; 16S RNA;  
 KW pyrophosphate dependent phosphofructokinase; nitrogen-containing compound;  
 KW ammonia; nitrate; nitrite; nitrogen; pigment; oxygen; landfill;  
 KW methane-containing environment; waste water treatment system; isoprenoid;  
 KW nitrous oxide; terpenoid; animal feed; carotenoid; exopolysaccharide.

OS Methylomonas sp.  
 OS  
 PN WO200220728-A2.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 28-AUG-2001; 2001WO-US026827.  
 XX  
 PR 01-SEP-2000; 2000US-0229858P.  
 XX  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Koffas M, Odom JM, Schenzle A;  
 XX  
 XX WPI; 2002-454358/48.  
 DR  
 DR N-PSDB; ABX83265.  
 XX  
 XX New high growth methanotrophic bacterial strain, useful for producing

PT single cell proteins, grows on a C1 carbon substrate, and comprises a  
 PT functional gene encoding in Embden-Meyerhof carbon pathway.  
 XX  
 XX Claim 11; Page 147-148; 157pp; English.

CC The invention relates to a high growth methanotrophic bacterial strain,  
 CC which grows on a C1 carbon substrate e.g. methane and methanol, and  
 CC comprises a functional Embden-Meyerhof carbon flux pathway comprising a  
 CC gene coding a pyrophosphate dependent phosphofructokinase enzyme or a 16S  
 CC RNA. The bacterial strain is useful for the production of single cell  
 CC protein and for the bio-transformation of a nitrogen-containing compound,  
 CC e.g. ammonia, nitrate, nitrite or nitrogen. It is also useful for the  
 CC production of a feed product comprising a protein, carbohydrates and a  
 CC pigment and for reducing oxygen demand, for removing nitrates and  
 CC nitrates in methane-containing environments such as landfills, waste  
 CC water treatment systems or anywhere that methane, oxygen and nitrates are  
 CC present. The bacterial strain of the invention can be used as a  
 CC denitrifying agent for the conversion of nitrate or nitrite to nitrous  
 CC oxide with methane or methanol as a carbon source. It is also used in the  
 CC production of biomass including proteins, carbohydrates and a wide  
 CC variety of pigments (particularly for isoprenoid pigments for the purpose  
 CC of generating animal feeds), in production of terpenoid and carotenoid  
 CC compounds, useful as pigments and as monomers in polymeric materials and  
 CC in production of exopolysaccharides at high levels. Sequences ABG61551-  
 CC ABG61590 represent high growth methanotrophic bacterial strain proteins  
 CC of the invention. (Updated on 07-AUG-2003 to correct OS field.)  
 XX

SQ Sequence 297 AA;

Query Match 100.0%; Score 1510; DB 5; Length 297;  
 Best Local Similarity 100.0%; Pred. No. 7.3e-140;  
 Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKIKAYLVCCQERVERALDARLPAENILPQTLHOAMEYSVLNGGKTRPLLTATGQAL 60  
 DB 1 MSKIKAYLVCCQERVERALDARLPAENILPQTLHOAMEYSVLNGGKTRPLLTATGQAL 60  
 QY 61 GLPENVLDAPACAVEFIHVYSLIHDDLPAMNDLRRGKPTCHKAYDEATATLQAL 120  
 DB 61 GLPENVLDAPACAVEFIHVYSLIHDDLPAMNDLRRGKPTCHKAYDEATATLQAL 120  
 QY 121 LAPEVLNDPGITVDAPARLKMITALTRASGQGWGGQADLGSVGRKLTLPLENMHI 180  
 DB 121 LAPEVLNDPGITVDAPARLKMITALTRASGQGWGGQADLGSVGRKLTLPLENMHI 180  
 QY 181 HTGALIRASVNLAAALSKPDLDTCAKLDHVAKCIGLSFQVKDILDIADTATLQKTQ 240  
 DB 181 HTGALIRASVNLAAALSKPDLDTCAKLDHVAKCIGLSFQVKDILDIADTATLQKTQ 240  
 QY 241 GKIDNDKPTYPALLGMAGAKQAQELHEQAVESLTGFGSEADLLRELSLYIERTH 297  
 DB 241 GKIDNDKPTYPALLGMAGAKQAQELHEQAVESLTGFGSEADLLRELSLYIERTH 297

RESULT 3  
 AAU80331  
 ID AAU80331 standard; protein; 297 AA.

XX AAU80331;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 XX Methylomonas 16a ORF7 ispA protein sequence.  
 DE  
 XX  
 KW Isoprenoid biosynthetic enzyme; isoprenoid compound; feed additive;  
 KW keratenoid; pigment; flavour; fragrance; open reading frame 7; ORF7;  
 KW ispA; geranyltransferase/farnesyl diphosphate synthase enzyme.

OS Methylomonas sp.  
 OS  
 XX  
 XX WO200220733-A2.  
 XX  
 XX 14-MAR-2002.

XX 29-AUG-2001; 2001WO-US026852.  
 XX 01-SEP-2000; 2000US-0229907P.  
 XX (DUPO) DU PONT DE NEMOURS & CO E I.  
 XX Cheng Q, Koffas M, Norton KC, Odum JM, Picataggio SK;  
 XX Rouviere PE, Schenzle A, Tomb J;  
 XX WPI; 2002-383051/41.  
 XX N-PSDB; ABK50087.  
 XX Novel nucleic acid molecule encoding a isoprenoid biosynthetic enzyme,  
 XX isolated from *Methylobacterium* 16a, useful for the production of isoprenoid  
 XX compounds.  
 XX Claim 4; Page 76-77; 84pp; English.  
 XX The present invention relates to a new nucleic acid molecule encoding an  
 XX isoprenoid biosynthetic enzyme isolated from *Methylobacterium* 16a. The  
 XX invention is useful for obtaining a nucleic acid molecule encoding an  
 XX isoprenoid compound biosynthetic enzyme, and for the microbial production  
 XX of isoprenoid compounds. The molecules of the invention are also useful  
 XX for regulating isoprenoid biosynthesis in an organism and for producing  
 XX recombinant organisms for producing various isoprenoid compounds. The  
 XX nucleic acid is also useful for feed additive, for the production of  
 XX xeroderms and their derivatives, isoprenoid intermediates, and as pure  
 XX products useful as pigments, flavours and fragrances. The present amino  
 XX acid sequence represents the *Methylobacterium* 16a open reading frame 7 (ORF7)  
 XX isPA (geranyltransferase or farnesyl diphosphate synthase enzyme) protein  
 XX of the invention, as described above  
 XX Sequence 297 AA;  
 XX  
 XX Query Match 100.0%; Score 1510; DB 5; Length 297;  
 XX Best Local Similarity 100.0%; Pred. No. 7.3e-140;  
 XX Matches 297; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX 1 MSKUKAYITVCOEVEERALDARLPAENILPOTLHOAMRYSVLNGSKETRELLTYATQAL 60  
 XX 1 MSKUKAYITVCOEVEERALDARLPAENILPOTLHOAMRYSVLNGSKETRELLTYATQAL 60  
 XX 61 GLPENLDPACAVEFHIVYSLIHDDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQA 120  
 XX 61 GLPENLDPACAVEFHIVYSLIHDDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQA 120  
 XX 121 LAPEVANDPGITVDAPARLKMHTALTRASGQGMVGGQAIIDLSGVGRKLTLELENMHI 180  
 XX 121 LAPEVANDPGITVDAPARLKMHTALTRASGQGMVGGQAIIDLSGVGRKLTLELENMHI 180  
 XX 181 HKTGALIRASVNLAAALSKPDLDTCAKLDHYAKICGLSFQVKKDILDEADTATLGKTQ 240  
 XX 181 HKTGALIRASVNLAAALSKPDLDTCAKLDHYAKICGLSFQVKKDILDEADTATLGKTQ 240  
 XX 241 GKDIDNKPTYPALLGMAGAKQKAEHQVSLTGFSGEADLLRELSLYIERTH 297  
 XX 241 GKDIDNKPTYPALLGMAGAKQKAEHQVSLTGFSGEADLLRELSLYIERTH 297  
 XX  
 XX RESULT 4  
 XX AAU36378  
 XX AAU36378 standard; protein; 295 AA.  
 XX AAU36378;  
 XX 14-FEB-2002 (first entry)  
 XX Pseudomonas aeruginosa cellular proliferation protein #368.  
 XX Antisense; prokaryotic cellular proliferation protein; antibiotic;  
 XX antibacterial; drug design.

OS Pseudomonas aeruginosa.  
 XX WO200170955-A2.  
 XX 27-SEP-2001.  
 XX 21-MAR-2001; 2001WO-US009180.  
 XX 21-MAR-2000; 2000US-0191078P.  
 XX 23-MAY-2000; 2000US-0206848P.  
 XX 26-MAY-2000; 2000US-0207727P.  
 XX 23-OCT-2000; 2000US-0242578P.  
 XX 27-NOV-2000; 2000US-0253625P.  
 XX 22-DEC-2000; 2000US-0257931P.  
 XX 16-FEB-2001; 2001US-0269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 XX Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 XX N-PSDB; AAS54237.  
 XX New polynucleotides for the identification and development of  
 XX antibiotics, comprise sequences of antisense nucleic acids.  
 XX Example 3; SEQ ID NO 11971; 51pp; English.  
 XX The invention relates to antisense inhibitors of genes essential to  
 XX prokaryotic cellular proliferation, their use in identifying the genes,  
 XX their use in the discovery of novel antibiotics, the essential genes,  
 XX themselves and the encoded proteins. The prokaryotes used are *Escherichia*  
 XX coli, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,  
 XX *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also  
 XX useful for the identification of potential new targets for antibiotic  
 XX development. The antisense nucleic acids can also be used to identify  
 XX proteins used in proliferation, to express these proteins, and to obtain  
 XX antibodies capable of binding to the expressed proteins. The proteins can  
 XX be used to screen compounds in rational drug discovery programmes. The  
 XX antisense nucleic acid sequence is also useful to screen for homologous  
 XX nucleic acids which are required for cell proliferation in a wide variety  
 XX of organisms. The present sequence represents an essential prokaryotic  
 XX cellular proliferation protein. Note: The sequence data for this patent  
 XX did not form part of the printed specification, but was obtained in  
 XX electronic format directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 295 AA;  
 XX  
 XX Query Match 55.0%; Score 831; DB 4; Length 295;  
 XX Best Local Similarity 60.3%; Pred. No. 4.7e-73;  
 XX Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;  
 XX  
 XX 4 LKAYITVCOEVEERALDARLPAENILPOTLHOAMRYSVLNGSKETRELLTYATQALGLP 63  
 XX 2 IAAVQARCOARVDAALDALFVAPREELQRYEAMRYSGMGKRVRLAYAAACEALGGA 61  
 XX 64 ENVLDPACAVEFHIVYSLIHDDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQAALF 123  
 XX 62 PQADAAACAVELIHAYSLVHDDLPAMDNDLRRGKPTCKHAYDEATAILAGDALQAALF 121  
 XX 124 EVLANDPGITVDAPARLKMHTALTRASGQGMVGGQAIIDLSGVGRKLTLELENMHIKT 183  
 XX 122 EVLANDTRNPQEHAVCLEMLTRARAGSAGMVGQAIIDLSGVGVALDQALAEVYHSHKT 181  
 XX 184 GALIRASVNLAAALSKPDLDTCAKLDHYAKICGLSFQVKKDILDEADTATLGKTQGD 243  
 XX 182 GALIRASVNLAAALSKPDLDTCAKLDHYAKICGLSFQVKKDILDEADTATLGKTQGD 241  
 XX 244 IDNKPTYPALLGMAGAKQKAEHQVSLTGFSGEADLLRELSLYIERTH 295  
 XX 242 QAHNKPTYPALLGMAGAKQKAEHQVSLTGFSGEADLLRELSLYIERTH 293

XX SQ Sequence 295 AA;  
Query Match 55.0%; Score 831; DB 6; Length 295;  
Best local Similarity 60.3%; Pred. NO. 4.7e-73;  
Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;  
XX AC ABU38665;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #24192.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Pseudomonas aeruginosa.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US0009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 23-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR N-PSDB; ACA42535.  
XX DR WPI; 2003-029926/02.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 25; SEQ ID NO 66589; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 6213 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
XX CC encoding a polypeptide whose expression is inhibited by the antisense  
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX CC polypeptide or its fragment whose expression is inhibited by the  
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX CC proliferation or the activity of a gene in an operon required for  
XX CC proliferation; (7) identifying a compound that influences the activity of  
XX CC the gene product or that has an activity against a biological pathway  
XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
XX CC identifying a gene required for cellular proliferation or the biological  
XX CC pathway in which a proliferation-required gene or its gene product lies  
XX CC or a gene on which the test compound that inhibits proliferation of an  
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX CC compound's activity; (11) a culture comprising strains in which the gene  
XX CC product is overexpressed or underexpressed; (12) determining the extent  
XX CC to which each of the strains is present in a culture or collection of  
XX CC strains; or (13) identifying the target of a compound that inhibits the  
XX CC proliferation of an organism. The antisense nucleic acids are useful for  
XX CC identifying proteins or screening for homologous nucleic acids required  
XX CC for cellular proliferation to isolate candidate molecules for rational  
XX CC drug discovery programs, or for screening homologous nucleic acids  
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX CC *X. pneumonae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX CC the target prokaryotic essential genes. Note: The sequence data for this  
XX CC patent did not form part of the printed specification, but was obtained  
XX CC in electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 295 AA;  
Query Match 55.0%; Score 831; DB 6; Length 295;  
Best local Similarity 60.3%; Pred. NO. 4.7e-73;  
Matches 176; Conservative 33; Mismatches 83; Indels 0; Gaps 0;  
QY 4 LKAYLVQCBRRVERALDARLPARNILPQTLHOAMRYSVLNGKRTPLITYATGQALGP 63  
DB 2 IAAVQACQARVDAALDALFVAPREELQRLYEAMRYSVNGKRVPLLAYAACALGGA 61  
QY 64 ENVIDAPACAVEPIHYSLIHDDLPAKNDLARGKPTCKHAYDEATAILAGDALQALAF 123  
DB 62 PQADAAACAVELIHAYSLVHDDLPAKNDLARGKPTCKHAYDEATAILAGDALQALAF 121  
QY 124 EVLADPGITVDAPARKMTALTTRASSQGMVGGOAIDLGSGVGRKLTLPLENNHHT 183  
DB 122 EVLADTRNPQEHAVCLEMLTLARAGSAGMVGGOAIDLGSGVALDQAALVWHERHT 181  
QY 184 GALIRASVNLALSKPDLDTCVAKLDHYAKICIGLSFQVKDILDEATATLTKTQGD 243  
DB 182 GALTEASVRLGALASGRABPASLAALERYAEATGLAFQVQDDILOVESDTATLTKTQGD 241  
QY 244 IDNDKPTYPALLGMAKAKOKAOLHEQAVESLTGSGSEADLLRELSLYIER 295  
DB 242 QAHNKPTYPALLGLEAKGVLELRDLALALDGGPPPSADPLQLARYIVER 293  
RESULT 6  
ABU40141  
ID ABU40141 standard; protein; 295 AA.  
XX AC ABU40141;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #25568.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Pseudomonas putida.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PR 21-MAR-2002; 2002WO-US0009107.  
XX PR 06-SEP-2001; 2001US-00815242.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR N-PSDB; ACA44011.  
XX DR WPI; 2003-029926/02.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 25; SEQ ID NO 68065; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 6213 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
XX CC encoding a polypeptide whose expression is inhibited by the antisense  
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX CC polypeptide or its fragment whose expression is inhibited by the  
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX CC proliferation or the activity of a gene in an operon required for  
XX CC proliferation; (7) identifying a compound that influences the activity of  
XX CC the gene product or that has an activity against a biological pathway  
XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
XX CC identifying a gene required for cellular proliferation or the biological  
XX CC pathway in which a proliferation-required gene or its gene product lies  
XX CC or a gene on which the test compound that inhibits proliferation of an  
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX CC compound's activity; (11) a culture comprising strains in which the gene  
XX CC product is overexpressed or underexpressed; (12) determining the extent  
XX CC to which each of the strains is present in a culture or collection of  
XX CC strains; or (13) identifying the target of a compound that inhibits the  
XX CC proliferation of an organism. The antisense nucleic acids are useful for  
XX CC identifying proteins or screening for homologous nucleic acids required  
XX CC for cellular proliferation to isolate candidate molecules for rational  
XX CC drug discovery programs, or for screening homologous nucleic acids  
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX CC *X. pneumonae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX CC the target prokaryotic essential genes. Note: The sequence data for this  
XX CC patent did not form part of the printed specification, but was obtained  
XX CC in electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 PR polypeptide or its fragment whose expression is inhibited by the  
 PR antisense nucleic acid; (4) an antibody capable of specifically binding  
 PR the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 PR proliferation or the activity of a gene in an operon required for  
 PR the gene product or that has an activity against a biological pathway  
 XX required for proliferation, or that inhibits cellular proliferation; (8)  
 PA identifying a gene required for cellular proliferation or the biological  
 XX pathway in which a proliferation-required gene or its gene product lies  
 PI or a gene on which the test compound that inhibits proliferation of an  
 PI organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 XX compound's activity; (11) a culture comprising strains in which the gene  
 XX product is overexpressed or underexpressed; (12) determining the extent  
 XX to which each of the strains is present in a culture or collection of  
 XX strains; or (13) identifying the target of a compound that inhibits the  
 XX proliferation of an organism. The antisense nucleic acids are useful for  
 XX identifying proteins or screening for homologous nucleic acids required  
 XX for cellular proliferation to isolate candidate molecules for rational  
 XX drug discovery programs, or for screening homologous nucleic acids  
 XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 XX the target prokaryotic essential genes. Note: The sequence data for this  
 XX patent did not form part of the printed specification, but was obtained  
 XX in electronic format directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 295 AA;

Query Match 53.7%; Score 811; DB 6; Length 295;  
 Best Local Similarity 58.8%; Pred. No. 4.3e-71;  
 Matches 171; Conservative 36; Mismatches 80; Indels 4; Gaps 3;  
 Y 7 YLTVQVERALDARLPAENILPOTLHQAMRYSVLNGGKTRPLITYATGQALGPENV 66  
 b 5 YQASQCARVDALEPLFIAPSKELRYAAMRYSVNGGKVRPLLAYAACEALGAPAEQ 64  
 Y 67 LDAPACAVEIHVYSLIHDDLPANDNDLRRGKPTCHKAYDEATAILAGDALQALAFVL 126  
 b 65 ANGAACAVELIHAYSLVHDDLPANDDDLLRRGQPTTHKAFDEACAILAGDGLQSLAF 124  
 Y 127 ANDPGIT--VDAPARKMHTALTFRASGQVMVGQAIDLGSGVGRKLTLPLENNMIHKTG 184  
 b 125 L-DPLSLQIDS-IRLAWQVLAKAAGPAGWVGQAIDLGSGVKLDQQALEFMRHKTG 182  
 Y 185 ALTRASVNLAALEKPDLTCAKLDHYAKCTGLSPQVQDDILDTEADTATLTKGTQKDI 244  
 b 183 ALIEASVRLGALASARAEPVQLDALQTYAQALGAFQVQDDILDVESDTATLTKGTQADI 242  
 Y 245 DNDKPTYPALLGMAGAKQAOLHEQAVESLTGFGSEADLLRELSLYIER 295  
 b 243 ARDKPTYPALLGLEAKAKAYAIERDQALAAALQGFGEKAPLALARYIVER 293

RESULT 7  
 BU41505  
 D ABU41505 standard; protein; 295 AA.  
 X ABU41505;  
 X ABU41505;  
 X 19-JUN-2003 (first entry)  
 X Protein encoded by prokaryotic essential gene #27032.  
 X Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 X Pseudomonas syringae.  
 X WO200277183-A2.  
 X 03-OCT-2002.  
 X 21-MAR-2002; 2002WO-US0009107.

XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 XX 06-MAR-2002; 2002US-0362699P.  
 PA (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
 XX WP7; 2003-029926/02.  
 XX N-PSDB; ACA45375.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 69429; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 295 AA;

Query Match 52.4%; Score 791; DB 6; Length 295;  
 Best Local Similarity 56.8%; Pred. No. 4e-69;  
 Matches 167; Conservative 36; Mismatches 91; Indels 0; Gaps 0;  
 QY 4 LKAYLTVQVERALDARLPAENILPOTLHQAMRYSVLNGGKTRPLITYATGQALGP 63  
 Db 2 IASYQSQSOTRVNAALEGLFQAPSPELSRLYAMRYSVNGGKVRPLLAYAACEALGV 61  
 QY 64 ENVLDPACAVEIHVYSLIHDDLPANDNDLRRGKPTCHKAYDEATAILAGDALQALAF 123  
 Db 62 AEDANGACAVELIHAYSLVHDDLPANDDDLLRRGQPTTHKAFDEACAILAGDGLQSLAF 121  
 QY 124 EVLANDPGITVDAPARKMHTALTFRASGQVMVGQAIDLGSGVGRKLTLPLENNMIHKT 183  
 Db 122 TALLAPHLSSRNAATRLQMVSTLAAAGPAGWVGQAIDLGSGVGLKLDQAALSYERHKT 181  
 QY 184 GALTIRASVNLAALEKPDLTCAKLDHYAKCTGLSPQVQDDILDTEADTATLTKGTQKDI 243

Db 182 GALIEASVRLGALSGQADQRLDALQVYRAVGLAFQVQDDILDVESDTATLGKRGAD 241  
QY 244 INDNKPVTYVLLGMLEAARSYAHLKVAEAVALLBPFQDKALELRQLAEFAVARKY 298  
Db 242 IARDKPTTYVLLGMLEAARSYAHLKVAEAVALLBPFQDKALELRQLAEFAVARKY 298  
RESULT 8  
ID ABP78160 standard; protein; 298 AA.  
XX AC ABP78160;  
XX DT 07-MAR-2003 (first entry)  
XX DE N. gonorrhoeae amino acid sequence SEQ ID 2850.  
XX KW Antibacterial; infection; vaccine; gene therapy.  
XX OS Neisseria gonorrhoeae.  
XX PN WO200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB002069.  
XX PR 12-FEB-2001; 2001GB-00003424.  
XX PA (CHIR-) CHIRON SPA.  
XX PI Pontana MR, Pizza M, Masignani V, Monaci E;  
XX WPI; 2003-058415/05.  
XX DR N-PSDB; AB241080.  
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
XX PT medicament for treating or preventing N. gonorrhoeae infection.  
XX PS Disclosure; Page 396; 815pp; English.  
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.  
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and  
XX CC antibodies that specifically bind to the proteins. The composition  
XX CC comprising the protein, nucleic acid or antibody is useful for the  
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
XX CC infection, this may be in the form of a vaccine or gene therapy.  
XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
XX CC molecules of the invention  
XX SQ Sequence 298 AA;  
Query Match 48.5%; Score 733; DB 6; Length 298;  
Best Local Similarity 51.7%; Pred. No. 2.1e-63;  
Matches 152; Conservative 44; Mismatches 96; Indels 2; Gaps 1;  
QY 4 LKAYLTVCQERVALDARLPALNIPOTLHQAMRYSVLNGGKTRPLLTATGQALGLP 63  
Db 7 LKAWQRAQAQTELLERFLPSGNEIPHTLHEAMRYAALDGGKLRPMLVLAASELGAM 66  
QY 64 ENVLDPACAVEFIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 123  
Db 67 ADAGQAAVAATIMIHVSLVHDDMPAMDNDLRRGKPTCHKIYGEATALLTGDAQTQAF 126  
QY 124 EYLANDPGITVDAPARLKMITALTASQSGQVWGQALDLSGVGRKLTLPLENWHIHT 183  
Db 127 DVLSP--TELPAAQLAMLSVLAAGSAGMAGQALDLANVGKQWQADLERHSLKT 184  
QY 184 GALIRASVNLAAALSKPDLDTCAVKKLDHYAKICIGLSFQVKDDILDIEADTATLTKTGCKD 243  
Db 185 GALIRAAVLGATACPDLSDAELAVDAYAAKLGAFQVDDVLDCEADTATLTKTAGKD 244  
QY 244 INDNKPVTYVLLGMLEAARSYAHLKVAEAVALLBPFQDKALELRQLAEFAVARKY 297

Db 245 ADNDKPTTYVLLGMLEAARSYAHLKVAEAVALLBPFQDKALELRQLAEFAVARKY 298  
RESULT 9  
ID ABP80110 standard; protein; 298 AA.  
XX AC ABP80110;  
XX DT 07-MAR-2003 (first entry)  
XX DE N. gonorrhoeae amino acid sequence SEQ ID 6750.  
XX KW Antibacterial; infection; vaccine; gene therapy.  
XX OS Neisseria gonorrhoeae.  
XX PN WO200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB002069.  
XX PR 12-FEB-2001; 2001GB-00003424.  
XX PA (CHIR-) CHIRON SPA.  
XX PI Pontana MR, Pizza M, Masignani V, Monaci E;  
XX WPI; 2003-058415/05.  
XX DR N-PSDB; AB241080.  
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
XX PT medicament for treating or preventing N. gonorrhoeae infection.  
XX PS Disclosure; Page 678; 815pp; English.  
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.  
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and  
XX CC antibodies that specifically bind to the proteins. The composition  
XX CC comprising the protein, nucleic acid or antibody is useful for the  
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
XX CC infection, this may be in the form of a vaccine or gene therapy.  
XX CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
XX CC molecules of the invention  
XX SQ Sequence 298 AA;  
Query Match 48.5%; Score 733; DB 6; Length 298;  
Best Local Similarity 51.7%; Pred. No. 2.1e-63;  
Matches 152; Conservative 44; Mismatches 96; Indels 2; Gaps 1;  
QY 4 LKAYLTVCQERVALDARLPALNIPOTLHQAMRYSVLNGGKTRPLLTATGQALGLP 63  
Db 7 LKAWQRAQAQTELLERFLPSGNEIPHTLHEAMRYAALDGGKLRPMLVLAASELGAM 66  
QY 64 ENVLDPACAVEFIHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 123  
Db 67 ADAGQAAVAATIMIHVSLVHDDMPAMDNDLRRGKPTCHKIYGEATALLTGDAQTQAF 126  
QY 124 EYLANDPGITVDAPARLKMITALTASQSGQVWGQALDLSGVGRKLTLPLENWHIHT 183  
Db 127 DVLSP--TELPAAQLAMLSVLAAGSAGMAGQALDLANVGKQWQADLERHSLKT 184  
QY 184 GALIRASVNLAAALSKPDLDTCAVKKLDHYAKICIGLSFQVKDDILDIEADTATLTKTGCKD 243  
Db 185 GALIRAAVLGATACPDLSDAELAVDAYAAKLGAFQVDDVLDCEADTATLTKTAGKD 244  
QY 244 INDNKPVTYVLLGMLEAARSYAHLKVAEAVALLBPFQDKALELRQLAEFAVARKY 297  
Db 245 ADNDKPTTYVLLGMLEAARSYAHLKVAEAVALLBPFQDKALELRQLAEFAVARKY 298

CC	standardise OS field)
XX	
SQ	Sequence 298 AA;
	Query Match 48.5%; Score 733; DB 6; Length 298;
	Best Local Similarity 51.7%; Pred. No. 2.le-63;
	Matches 152; Conservative 44; Mismatches 96; Indels 2; Gaps 1;
QY	4 LKAVLTVQCRVERPDAALPAENILPQLTHQAMYSVLNGSKTRPLTATGQALGLP 63     : : : : :     : : : : :     : : : : :     : : : : :
DB	7 LKAQQQAQAQTLLERFLPSNETPHTLHEAMYAALDGGKRURPEMLVLAASELGAM 66     : : : : :     : : : : :     : : : : :     : : : : :
QY	64 ENVLADAPACAVEITHVYSLIHDDLPAMDNDLRRGKPTCKEAYDEATAIAGDALQALAF 123     : : : : :     : : : : :     : : : : :     : : : : :
DB	67 ADAYGQARAIEMHIVYSLVHDDMPAMDNSLRREGKPTCHI KYGEATALLTGDALETOAF 126     : : : : :     : : : : :     : : : : :     : : : : :
QY	124 EVLANDPGITVDAPARKMITALTTRASSQGVMVGQAIDLSGVGKLTLPLENNHIHT 183     : : : : :     : : : : :     : : : : :     : : : : :
DB	127 DVLGRP--TELPAAQOLAMLVLAKAGSGSAGWAGQAILANVGKQMVQADLERWESLKT 184     : : : : :     : : : : :     : : : : :     : : : : :
QY	184 GALLRASVNLAAALSKEPDLOTCTCAKLHDHYAKICIGLSFQVKDDIIDIEADTATLGKTKGD 243     : : : : :     : : : : :     : : : : :     : : : : :
DB	185 GALLRAAVILGATACEDLSDAELAVLDAYAACLGLAFQVIDDVLDCEADTATLGKTAGKD 244     : : : : :     : : : : :     : : : : :     : : : : :
QY	244 IDNKPYPYLLGAGAKOKAOBELHQAVESLTGFGSEADILLRELUSLYIIERTH 297       : : : : :     : : : : :     : : : : :     : : : : :
DB	245 ADNDKPTVVYLMGLEARSYAHLVAEAVALLEPFQDKALRLQLAEFAVARKY 298       : : : : :     : : : : :     : : : : :     : : : : :
RESULT 11	
ABU38181	ID ABU38181 standard; protein; 298 AA.
XX	
AC	ABU38181;
DT	19-JUN-2003 (first entry)
DE	Protein encoded by Prokaryotic essential gene #23708.
KW	Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS	Neisseria meningitidis.
XX	
PN	WO200277183-A2.
PD	03-OCT-2002.
PF	
PP	21-MAR-2002; 2002WO-US009107.
PR	21-MAR-2001; 2001US-00815242.
PR	06-SEP-2001; 2001US-00948993.
PR	25-OCT-2001; 2001US-0342923P.
PR	08-FEB-2002; 2002US-00072851.
PR	06-MAR-2002; 2002US-0362699P.
XX	(ELIT-) ELITRA PHARM INC.
XX	
XX	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI	Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
DR	
XX	WPI; 2003-029926/02.
XX	N-PSDB; ACA42051.
DR	
DR	New antisense nucleic acids, useful for identifying proteins or screening
PT	for homologous nucleic acids required for cellular proliferation to
PT	isolate candidate molecules for rational drug discovery programs.
XX	
XX	Claim 25; SEQ ID NO 66105; 1766pp; English.
CC	The invention relates to an isolated nucleic acid comprising any one of
CC	the 6213 antisense sequences given in the specification where expression
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:
CC	(1) a vector comprising a promoter operably linked to the nucleic acid



CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 298 AA;

Query Match 48.3%; Score 729; DB 6; Length 298;  
 Best Local Similarity 51.4%; Pred. No. 5.2e-63;  
 Matches 151; Conservative 44; Mismatches 97; Indels 2; Gaps 1;  
 QY 4 LKAYLVQCEVEREALDARLPAENILPOTLHQAMRYSVLNGGKRTPLTYATGALGLP 63  
 DB 7 LKAWQRAQAQOTELLERFLPSENEIPHTLHEAMRYAALDGGKRLPMLVLAASLGERAV 66  
 QY 64 ENVLDPACAVEFTHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 123  
 DB 67 HEANGQAWAA-EMIHASLVHDDPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAF 126  
 QY 124 EVLANDPOITVDAPARLKMITALTRASSGQGWGGQALDLSVGRKLTPELENHHIKT 183  
 DB 127 DVLGRP--TELPAAQQLAMLSVLAKAGSGMGAGQALDLSVGRKLTPELENHHIKT 184  
 QY 184 GALLIRASVNLAAKSPDLDTVCVAKKLDHYAKCIGLSFQVKDDILDLEADTATLGTKGKD 243  
 DB 185 GALLIRAAVLLGAMSCPNLSDESLAVLDVYAKLGLAFQVDDVLDCEADTATLGTKGKD 244  
 QY 244 IDNDKPTVYDALLGAGAKQKQAEHLHQAVESLTFGSGEADLLRELSLYIERTH 297  
 DB 245 ADNDKPTVYKLMGLEARSYAKHLVAEVALLEPFGDKALRLQLAEFAVARY 298

RESULT 12

ID ASU21390  
 XX ID ABU21390 standard; protein; 293 AA.  
 XX AC ABU21390;

DT 19-JUN-2003 (first entry)

DE Protein encoded by prokaryotic essential gene #5917.

DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

KW Burkholderia fungorum.

OS W020027183-A2.

PN 03-OCT-2002.

XX

PF 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 28-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0363699P.  
 XX (BLIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PR Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA25260.  
 DR New antisense nucleic acids, useful for identifying proteins or screening  
 PR for homologous nucleic acids required for cellular proliferation to  
 PR isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 49314; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 293 AA;

Query Match 47.6%; Score 719.5; DB 6; Length 293;  
 Best Local Similarity 53.2%; Pred. No. 4.3e-62;  
 Matches 151; Conservative 38; Mismatches 90; Indels 5; Gaps 2;

QY 13 ERVERALDARLPAENILPOTLHQAMRYSVLNGGKRTPLTYATGALGLPENVLDPAC 72  
 DB 12 ERVETALDHYLFGTEATEPATLHEAMRYAVLGGKRVPLCHRAAGELTGARAECLDAAA 71  
 QY 73 AVEFTHVYSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAEVLNDPGI 132  
 DB 72 ALEHVVYSLVHDDMPCMDLRRGKPTCHKAYDEATAILAGDALQALAEVLNDPGI 128  
 QY 133 TVDAPA-RLKMITALTRASSGQGWGGQALDLSVGRKLTPELENHHIKTALIRASV 191  
 DB 129 -VLAPAAQAALVRELALASSGQGWGGQALDLSVGHGTLTFTQLETHWMTKALLRAV 187  
 QY 192 NLAALS KPDLDTCVAKKLDHYAKCIGLSFQVKDDILDLEADTATLGTKGKDINDKPT 251

188 RMGALAGETPDAAAMRSLDAYSAAVGLAFQVDDILDVTTDSATLGTAKGAKGKPTY 247  
252 PALLGMAGAKOKQELHEQAVESITGFGSEADLLRELSLYIIEER 295  
248 VSIIGLDASRAALAAQLSDHAHALAPFGAQAQRLAELADLVNR 291

## RESULT 13

ABU19741  
D ABU19741 standard; protein; 294 AA.

ABU19741;

19-JUN-2003 (first entry)

Protein encoded by prokaryotic essential gene #5268.

Antisense; prokaryotic essential gene; cell proliferation; drug design.

Borrelia cepacia.

WO200277183-A2.

03-OCT-2002.

21-MAR-2002; 2002WO-US009107.

21-MAR-2001; 2001US-00815242.

06-SEP-2001; 2001US-00948993.

25-OCT-2001; 2001US-0342923P.

08-FEB-2002; 2002US-00072851.

06-MAR-2002; 2002US-0362699P.

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

WPI; 2003-029926/02.

N-PSDB; ACA23611.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 47665; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids

CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 294 AA;

Query Match 47.1%; Score 711; DB 6; Length 294;

Best Local Similarity 53.0%; Pred. No. 3e-61;

Matches 151; Conservative 41; Mismatches 87; Indels 6; Gaps 3;

QY 13 ERVERALDARLPARNILPQTLHOAMRYSVLNGKRTPLITYATGQALGUPENVLDAPAC 72

DB 12 DRVEDALGHYLPAAETAMPKLEHMYAVLVGGKRVRLPCHAAAGELTGATEAARNA 71

QY 73 AVBFHYVSLIHDDLPAMDNDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGI 132

DB 72 ALEMHVYSLVHDDMPDMDLRRGKPTCHKAYDEATAILAGDALQALAFVLANDPGI 131

QY 133 TVDAPAELEKMITALTRASGQGVGGQIDLSVGRKLTPELENMHHKGTALIRASVN 192

DB 132 SPVQQAAL--VRELALASGSGIAGGQIDLSVGLKLTREQLTMRMKTGALLRAAVR 189

QY 193 LAALS--KPOLDTCAKKLDHYAKCIGLSFQVDDILDIEADTATLGTCKGKDIDNDKPT 250

DB 190 MGALAGETPTSTTMAA--LDVYAGAVGLAFQVDDILDVTTDSATLGTAKGAKDNDKPT 247

QY 251 YPALLGMAGAKOKQELHEQAVESITGFGSEADLLRELSLYIIEER 295

DB 248 YVSLIGLEASRELAQAQLRAEADHAKPFPGARAQRLAELADLVNR 292

## RESULT 14

ABU28056

ID ABU28056 standard; protein; 299 AA.

AC ABU28056;

DT 19-JUN-2003 (first entry)

DE Protein encoded by prokaryotic essential gene #13583.

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

CS Enterobacter cloacae.

FN WO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DE N-PSDB; ACA31926.

XX New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 55980; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 299 AA;

Query Match 47.1%; Score 711; DB 6; Length 299;  
Best Local Similarity 54.1%; Pred. No. 3.1e-61;  
Matches 158; Conservative 35; Mismatches 89; Indels 10; Gaps 3;

QY 12 QERVERALDARPAENILPQ-----TLHQMYRYSVNGKTRPLLTATGALGLPENV 66  
DB 8 QARVWVANDAL--RRFAPQPTNTPLVAMHYGALLGKRLRPPLVATGNNFGISNT 65  
QY 67 LDAPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHXAYDEATAILAGDALQALAEVL 126  
DB 66 LDAPAAVECHAYSLIHDDLPAMDNDLRRGKPTCHXAYDEATAILAGDALQALAEVL 125  
QY 127 ANDPGITVDAPARLKMITALTRASSGQMGVGGQALDLSGVGKRLTLPLENMHIHKTAL 186  
DB 126 SDAPMAEVAADRORLAMISELAWASGVAGMCGGQALDLAEAGRVNLEQLERHHRKTAL 185  
QY 187 IRASVNLAAALSKPDLTVCVAKKLDHYAKICIGLSPOVKDDILDIEADTATLGTGKQIDN 246  
DB 186 IRSVAVLGLSAGEQGRKALPILDRYASISGLAFQVQDDILDVVGDTATLGRQADQOL 245  
QY 247 DKPTYPALLGMAGAKQAKOELHQAQVESLTGFGSEA---DLRLSLYIIER 295  
DB 246 GKSTYPALLGLEAQAKARDLIDARQSLNELAAQSLDLSALEALADYIIOR 297

RESULT 15

AAU38491  
ID AAU38491 standard; protein; 299 AA.

XX AC AAU38491;  
XX DT 14-FEB-2002 (first entry)  
XX DE Salmonella typhi cellular proliferation protein #382.  
XX AN Antisense; prokaryotic cellular proliferation protein; antibiotic;  
XX KW antibacterial; drug design.

OS Salmonella typhi.  
XX PN WO200170955-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US009180.  
XX PR 21-MAR-2000; 2000US-0191078P.  
XX PR 23-MAY-2000; 2000US-0206848P.  
XX PR 26-MAY-2000; 2000US-0207727P.  
XX PR 23-OCT-2000; 2000US-0242578P.  
XX PR 21-NOV-2000; 2000US-0253625P.  
XX PR 22-DEC-2000; 2000US-0257931P.  
XX PR 16-FEB-2001; 2001US-0269308P.  
XX PA (BLIT-) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu RH;  
XX DR WPI; 2001-611495/70.  
XX DR N-PSDB; AAS56350.  
XX PT New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids.

XX Example 3; SEQ ID NO 14084; Slipp; English.

The invention relates to antisense inhibitors of genes essential to  
prokaryotic cellular proliferation, their use in identifying the genes,  
their use in the discovery of novel antibiotics, the essential genes  
themselves and the encoded proteins. The prokaryotes used are *Escherichia*  
*coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,  
*Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also  
useful for the identification of potential new targets for antibiotic  
development. The antisense nucleic acids can also be used to identify  
proteins used in proliferation, to express these proteins, and to obtain  
antibodies capable of binding to the expressed proteins. The proteins can  
be used to screen compounds in rational drug discovery programmes. The  
antisense nucleic acid sequence is also useful to screen for homologous  
nucleic acids which are required for cell proliferation in a wide variety  
of organisms. The present sequence represents an essential prokaryotic  
cellular proliferation protein. Note: The sequence data for this patent  
did not form part of the printed specification, but was obtained in  
electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 299 AA;

Query Match 46.7%; Score 705.5; DB 4; Length 299;  
Best Local Similarity 52.7%; Pred. No. 1.1e-60;  
Matches 156; Conservative 39; Mismatches 94; Indels 7; Gaps 3;

QY 3 KLKAYLTVCQERVERALDARLPAENILPOTLHQMYRYSVNGKTRPLLTATGALGL 62  
DB 6 QLCQVYQANQALSRI-APLPFN---TPVVEMQYGALLGKRLRPPLVATGQMFV 61  
QY 63 PENVLDPACAVEFIHVYSLIHDDLPAMDNDLRRGKPTCHXAYDEATAILAGDALQALA 122  
DB 62 STATLDAPAAVECHAYSLIHDDLPAMDNDLRRGLPTCHIKFGEANAILAGDALQTLA 121  
QY 123 FEVLANDPGITVDAPARLKMITALTRASSGQMGVGGQALDLSGVGKRLTLPLENMHIH 182  
DB 122 FTIISDAPMEVADRDRIAMIAELANASGIAGMCGGQALDLAEGQRITLDALEIRHRK 181  
QY 183 TGLIRASVNLAAALSKPDLTVCVAKKLDHYAKICIGLSPOVKDDILDIEADTATLGTG 242  
DB 182 TGLIRASVNLAAALSKPDLTVCVAKKLDHYAKICIGLSPOVKDDILDIEADTATLGTG 241  
QY 243 DIDNDKPTYPALLGMAGAKQAKOELHQAQVESLTGFGSEA---DLRLSLYIIER 295  
DB 242 DQGLGKSTYPALLGLEAQAKARDLIDARQSLNELAAQSLDLSALEALADYIIOR 297

Search completed: February 29, 2004, 14:43:39  
Job time : 41.2152 secs